



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 100034

TO: Anish Gupta
Location: cm-1/9A13/11D13
Art Unit: 1654

July 30, 2003

Case Serial Number: 09/787070

From: P. Sheppard
Location: CM1-1E03
Phone: (703) 308-4499

sheppard@uspto.gov

Search Notes

THIS PAGE BLANK (USPTO)

STIC-Bi tech/Ch mLib

100034

From: Gupta, Anish
Sent: Tuesday, July 29, 2003 3:53 PM
To: STIC-Biotech/ChemLib
Subject: search request

Application Serial Number: 09/787,070

location: 9A13

Art Unit: 1654

Date: 7-27-03

phone: 308-4001

Please search seq. ID NO. 1-8

thank you
anish gupta

Point of Contact
P. Sheppard
Telephone number: (703) 308-4499

7/30/03

Point of Contact
P. Sheppard
Telephone number: (703) 308-4499

RECEIVED
JUL 29 2003
STIC

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 27.1429 Seconds
(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-1

Sequence: 1 VYOHQAKMPWIOPKT 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the local score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	81	88.0	210	6	O62825 bubalus bub
2	77	83.7	223	6	O9MYU6 capra hircu
3	68	73.9	223	6	O9TYU7 capra hircu
4	69	73.9	223	6	O9GK07 capra hircu
5	50	54.3	693	11	O9WV66 mus musculu
6	48	52.2	225	4	O9BTR9 O9BTR9 homo sapien
7	48	52.2	328	16	O8CP60 O8CP60 staphylococ
8	48	52.2	418	16	O9WV65 thermotoga
9	48	52.2	704	4	O9H992 O9H992 homo sapien
10	46.5	50.5	363	5	O01557 O01557 caenorhabdi
11	46.5	50.5	364	5	O01555 O01555 caenorhabdi
12	46	50.0	265	8	O9Z2R2 cyandidiosch
13	46	50.0	458	4	O8WVP9 homo sapien
14	46	50.0	475	4	O96FK5 homo sapien
15	46	50.0	893	4	O9COB8 O9COB8 homo sapien
16	45	48.9	309	10	O9FJP9 O9FJP9 arabidopsis

17	45	48.9	311	5	Q19257 Q19257 caenorhabdi
18	45	48.9	931	10	O8SB48 O8SB48 oryza sativ
19	44.5	48.4	205	13	O9PW95 O9PW95 anas platyr
20	44	47.8	80	10	O81534 O81534 saccharum o
21	44	47.8	85	10	O81399 O81399 orobanche c
22	44	47.8	328	17	O9UYA2 O9UYA2 pyrococcus
23	44	47.8	340	10	O96568 O96568 ipomoea pur
24	44	47.8	379	10	O04065 O04065 perilla fru
25	44	47.8	386	10	O23923 O23923 digitalis l
26	44	47.8	388	10	O9PS37 O9PS37 torenia hyb
27	44	47.8	389	10	O8LP19 O8LP19 niterbergi
28	44	47.8	389	10	O43040 O43040 petunia hyb
29	44	47.8	389	10	O933P8 O933P8 nicotiana t
30	44	47.8	389	10	O9MSB2 O9MSB2 petunia hyb
31	44	47.8	390	10	O9LRE2 O9LRE2 scutellaria
32	44	47.8	390	10	O48564 O48564 scutellaria
33	44	47.8	390	10	O22122 O22122 scutellaria
34	44	46.7	390	2	F74958 F74958 shewanella
35	43	46.7	69	5	O25047 O25047 haemochus
36	43	46.7	136	5	O25044 O25044 haemochus
37	43	46.7	161	5	O26266 O26266 caenorhabdi
38	43	46.7	180	5	O25045 O25045 haemochus
39	43	46.7	205	5	O9N6M8 O9N6M8 drosophila
40	43	46.7	279	8	O9G3M0 O9G3M0 arabidopsis
41	43	46.7	279	10	O8LB27 O8LB27 arabidopsis
42	43	46.7	279	10	O9LRT2 O9LRT2 arabidopsis
43	43	46.7	297	5	O44075 O44075 caenorhabdi
44	43	46.7	308	2	O9REU7 O9REU7 agrobacteri
45	43	46.7	393	11	O91XA2 O91XA2 mus musculu

ALIGNMENTS

RESULT 1					
ID	O62825	PRELIMINARY;	PRT;	210 AA.	
AC	O62825;				
DT	01-AUG-1998 (TREMBLrel. 07, Created)				
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)				
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)				
DE	AS2-casein (Fragment)				
OS	Bubalus bubalis (Domestic water buffalo)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bubalus.				
OX	NCBI_TaxID=99462;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
KC	TISSUE=Mammary gland;				
RA	Dae P., Jain S., Garg L.C.;				
RT	"Cloning and nucleotide sequence of cDNA encoding as2-casein in B.				
RT	bubalis."				
RL	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AJ005431; CAA06534.2; "				
DR	InterPro: IPR001588; Casein.				
DR	Pfam: PF00363; caseins; 2.				
FT	NON TER				
FT	SEQUENCE 210 AA; 24700 MW; 05DEF95963F1132C CRC64;				
QY					
Db	1 VYOHQAKMPWIOPKT 16				
	186 VYOHQAKMPWIOPKT 201				
Query Match	88.0%; Score 81; DB 6; Length 210;				
Best Local Similarity	87.5%; Pred. No. 1.5e-05;				
Matches	14; Conservative 1; Mismatches 1; Indels 0;				
RESULT 2					
O9MYU6	PRELIMINARY;	PRT;	223 AA.		
ID	O9MYU6				
AC	O9MYU6;				

```

DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Alpha s2-casein.
GN CSN152.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA Lagomigro R., Pilla F., Matassino D., Zullo A.;
RT "A new allele of goat alpha s2-casein gene."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ289716; CAB94236.1; -.
DR InterPro: IPR001588; Casein.
DR Pfam: PF00363; caseins; 2.
DR PROSITE: PS00306; CASEIN_ALPHA_BETA; 1.
SQ SEQUENCE 223 AA; 26403 MW; 0E1F83F24DA85E2 CRC64;

Query Match 83.7%; Score 77; DB 6; Length 223;
Best Local Similarity 87.5%; Pred. No. 7.5e-05;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPKT 16
DB 199 VDOHOKAMKPMIOPKT 214

RESULT 3
Q9TT07 PRELIMINARY; PRT; 223 AA.
ID Q9TT07;
AC Q9TT07;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Alpha s2-casein.
GN CSN152.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA Velti C.C., Pilla F., Lagomigro R.R.;
RT "A new allele of goat alpha s2-casein."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ249995; CAB59920.1; -.
DR InterPro: IPR001588; Casein.
DR Pfam: PF00363; caseins; 2.
DR PROSITE: PS00306; CASEIN_ALPHA_BETA; 1.
SQ SEQUENCE 223 AA; 26433 MW; CE9F4DC8D7688293 CRC64;

Query Match 73.9%; Score 68; DB 6; Length 223;
Best Local Similarity 81.2%; Pred. No. 0.0023;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPKT 16
DB 199 VDOHOKAMKPMIOPKT 214

RESULT 4
Q9GK07 PRELIMINARY; PRT; 223 AA.
ID Q9GK07;
AC Q9GK07;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Alpha s2-casein.
GN CSN152.

```

```

OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA Velti C.C., Pilla F., Lagomigro R.R.;
RT "A new allele of alpha s2-casein."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ287310; CAC21704.2; JOINED.
DR EMBL; AJ287311; CAC21704.2; JOINED.
DR EMBL; AJ242728; CAC21704.2; JOINED.
DR EMBL; AJ297312; CAC21704.2; JOINED.
DR EMBL; AJ297313; CAC21704.2; JOINED.
DR EMBL; AJ297314; CAC21704.2; JOINED.
DR EMBL; AJ242527; CAC21704.2; JOINED.
DR EMBL; AJ297315; CAC21704.2; JOINED.
DR EMBL; AJ287316; CAC21704.2; JOINED.
DR EMBL; AJ242526; CAC21704.2; JOINED.
DR EMBL; AJ242528; CAC21704.2; JOINED.
DR EMBL; AJ242533; CAC21704.2; JOINED.
DR InterPro: IPR001588; Casein.
DR Pfam: PF00363; caseins; 2.
DR PROSITE: PS00306; CASEIN_ALPHA_BETA; 1.
SQ SEQUENCE 223 AA; 26432 MW; CE9765E8D7688C9D CRC64;

Query Match 73.9%; Score 68; DB 6; Length 223;
Best Local Similarity 81.2%; Pred. No. 0.0023;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPKT 16
DB 199 VDOHOKAMKPMIOPKT 214

RESULT 5
Q9WV66 PRELIMINARY; PRT; 693 AA.
ID Q9WV66;
AC Q9WV66;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Axotrophin.
GN AXOT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Haendel M.A., Wagner M.K., Lyons G.E.;
RT "Insertional mutagenesis of a novel gene, axotrophin, causes callosal
RT agenesis and neural degeneration."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF155739; AA038411.1; -.
DR EMBL; BC025029; AA025029.1; -.
DR MGI:1931053; Axot.
SQ SEQUENCE 693 AA; 76598 MW; 7082766DE39E4CED CRC64;

Query Match 54.3%; Score 50; DB 11; Length 693;
Best Local Similarity 57.1%; Pred. No. 7;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 YOHOKAMKPMIOPK 15
DB 580 YVHOKAMKPMIOPK 593

```

RESULT 6

09BTR9 PRELIMINARY; PRT; 225 AA.

AC 09BTR9
 DT 01-JUN-2001 (TRENBLREL. 17, Created)
 DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)
 DT 01-OCT-2002 (TRENBLREL. 22, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Straussberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC003404; AA003404.1; -.
 KW Hypothetical protein.
 FT NON_TER
 SQ SEQUENCE 225 AA; 25142 MW; 261C78E4FFFA66E CRC64;

Query Match

52.2%; Score 48; DB 4; Length 225;
 Best Local Similarity 57.1%; Pred. No. 4.8;

Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 QY 2 YOHOKAKMPWIOPK 15
 DB 99 YVHODCKMKWLQAK 112

RESULT 7

08CP60 PRELIMINARY; PRT; 328 AA.

AC 08CP60
 DT 01-MAR-2003 (TRENBLREL. 23, Created)
 DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE Thioleodoxin reductase-like protein.
 GN SE1168.
 OS Staphylococcus epidermidis.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcaceae.
 OX NCBI_TaxID=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12228;
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z., Chen Z., Wen Y.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB016748; AA004767.1; -.
 KW Complete proteome.
 SQ SEQUENCE 328 AA; 36814 MW; 286F6B324EB6F887 CRC64;

Query Match

52.2%; Score 48; DB 16; Length 328;
 Best Local Similarity 66.7%; Pred. No. 7;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 3 OHOKAKMPWIOPK 14
 DB 190 QYPAIKPWILP 201

RESULT 8

09WYMS PRELIMINARY; PRT; 418 AA.

AC 09WYMS
 DT 01-NOV-1999 (TRENBLREL. 12, Created)
 DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE PHOH-related protein.
 GN TM0495.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.

OX NCBI_TaxID=2336;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.B., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A., McDonald L., Utterback T.R., Malek J.A., Lither K.D., Garrett M.M., Stewart A.M., Cotton M.D., Pratt W.S., Phillips C.A., Richardson D., Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O., Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima."
 RL Nature 399:323-329(1999).
 DR EMBL: AB001726; AAD35580.1; -.
 DR TIGR: TM0495; -.
 DR InterPro: IPR003714; PhOH.
 DR InterPro: IPR005596; PINC.
 DR Pfam: PF02562; PhOH; 1.
 DR SMART: SM00670; PINC; 1.
 KW Complete proteome.
 SQ SEQUENCE 418 AA; 47518 MW; A42EB74F86D7C8FB CRC64;

Query Match

52.2%; Score 48; DB 16; Length 418;
 Best Local Similarity 70.0%; Pred. No. 9;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 5 QKAKMPWIOPK 14
 DB 285 EKKKMKPWLOPK 294

RESULT 9

09H992 PRELIMINARY; PRT; 704 AA.

AC 09H992
 DT 01-MAR-2001 (TRENBLREL. 16, Created)
 DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)
 DT 01-OCT-2002 (TRENBLREL. 22, Last annotation update)
 DE Hypothetical protein FLJ12911.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Iwagaki T., Ota T., Hayashi K., Sugiyama T., Otauki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y., Matsumura H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Sasaki N.;
 RT "NEBO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK022973; BAB14340.1; -.
 DR Genew; HGNC:17393; AXOT.
 KW Hypothetical protein.
 SQ SEQUENCE 704 AA; 78050 MW; 1271F74D7F389130 CRC64;

Query Match

52.2%; Score 48; DB 4; Length 704;
 Best Local Similarity 57.1%; Pred. No. 15;

Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 QY 2 YOHOKAKMPWIOPK 15
 DB 578 YVHODCKMKWLQAK 591

RESULT 10

001557 PRELIMINARY; PRT; 363 AA.

ID 001557
 AC 001557
 DT 01-JUL-1997 (TRENBLREL. 04, Created)
 DT 01-JUL-1997 (TRENBLREL. 04, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)

DE F47B3.1 protein.
 GN F47B3.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 NCBI_TaxID=6239;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Favello A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaison N., Smith A., Sonnenhammer B., Staden K., Sulston J.,
 RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Spirot J., Wohldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Du Z., Le T.T.;
 RT "The sequence of C. elegans cosmid F47B3.";
 RT Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U97017; AAB52361.1; -.
 DR HSSP: P29350; IGMZ.
 DR WormPep: F47B3.1; CE10616.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR InterPro: IPR000242; TYR_PP.
 DR Pfam: PF00102; Y_phosphatase; 1.
 DR PRINTS: PR00700; PRYPHPTASE.
 DR SMART: SM00194; PTPC; 1.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
 DR HydroLase.
 SQ SEQUENCE 363 AA; 40796 MW; 58AF84204CBDD011.CRC64;
 Query Match 50.5%; Score 46.5; DB 5; Length 363;
 Best Local Similarity 69.2%; Pred. No. 14;
 Matches 9; Conservative 1; Mismatches 0; Indels 3; Gaps 1;
 QY 4 HOKAM---KPMIQ 13
 DB 48 HOKAMIVKPMVQ 60
 RESULT 11
 001555 PRELIMINARY; PRT; 364 AA.
 ID 001555
 AC 001555;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE F47B3.6 protein.
 GN F47B3.6
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 NCBI_TaxID=6239;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;

RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Favello A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaison N., Smith A., Sonnenhammer B., Staden K., Sulston J.,
 RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Spirot J., Wohldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Du Z., Le T.T.;
 RT "The sequence of C. elegans cosmid F47B3.";
 RT Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U97017; AAB52361.1; -.
 DR HSSP: P29350; IGMZ.
 DR WormPep: F47B3.1; CE10626.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR InterPro: IPR000242; TYR_PP.
 DR Pfam: PF00102; Y_phosphatase; 1.
 DR PRINTS: PR00700; PRYPHPTASE.
 DR SMART: SM00194; PTPC; 1.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
 DR HydroLase.
 SQ SEQUENCE 364 AA; 40860 MW; B2B779C1D43F44B8.CRC64;
 Query Match 50.5%; Score 46.5; DB 5; Length 364;
 Best Local Similarity 69.2%; Pred. No. 14;
 Matches 9; Conservative 1; Mismatches 0; Indels 3; Gaps 1;
 QY 4 HOKAM---KPMIQ 13
 DB 49 HOKAMIVKPMVQ 61
 RESULT 12
 092ZR2 PRELIMINARY; PRT; 265 AA.
 ID 092ZR2
 AC 092ZR2;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Succinate dehydrogenase from sulfur protein.
 GN SDHB
 OS Cyanidioschyzon merolae.
 OC Mitochondrion.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Cyanidioschyzon.
 NCBI_TaxID=45157;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=10D;
 RX MEDLINE=99030526; PubMed=9801310;
 RA Ohta N., Sato N., Kuroiwa T.;
 RT "Structure and organization of the mitochondrial genome of the
 RT unicellular red alga Cyanidioschyzon merolae deduced from the complete
 RT nucleotide sequence.";
 RL Nucleic Acids Res. 26:5190-5198(1998).
 CC -1- COPOLYMER BINDS 1 2FE-2S CLUSTER (BY SIMILARITY).
 DR EMBL: D89861; BAA34653.1; -.
 DR HSSP: P00364; IKF6.

DR InterPro; IPR006058; 2FE2S_ferredoxin.
DR InterPro; IPR001450; 4FE4S_ferredoxin.
DR InterPro; IPR004489; DhsB.
DR InterPro; IPR001041; Ferredoxin.
DR Pfam; PF00111; Fer2.1.
DR TIGRFAMs; TIGR00384; dhsB.1.
DR PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
KW Iron; Iron-sulfur; Mitochondrion.
SQ SEQUENCE 265 AA; 30924 MW; D3ED87DC55497D9 CRC64;

Query Match 50.0%; Score 46; DB 8; Length 265;
Best Local Similarity 46.2%; Pred. No. 12;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 YVOHQKMKPMIQ 14
Db 134 YTGKSTQPMIQ 146

RESULT 13

O8WVP9 PRELIMINARY; PRT; 458 AA.
AC O8WVP9;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strauberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017658; AAH17658.1; -
DR InterPro; IPR003659; Plexin-like.
DR InterPro; IPR002165; Plexin_repeat.
DR InterPro; IPR001627; Sema.
DR Pfam; PF01437; PSI; 1.
DR Pfam; PF01403; Sema; 1.
DR SMART; SM00423; PSI; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 458 AA; 50263 MW; B8A5233DCE848E7 CRC64;

Query Match 50.0%; Score 46; DB 4; Length 458;
Best Local Similarity 61.5%; Pred. No. 21;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 YVOHQKMKPMIQ 13
Db 176 LYOPQATRPWQ 188

RESULT 14

O96FK5 PRELIMINARY; PRT; 475 AA.
AC O96FK5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strauberg R.;

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010701; AAH10701.1; -
DR InterPro; IPR003659; Plexin-like.
DR InterPro; IPR002165; Plexin_repeat.
DR InterPro; IPR001627; Sema.
DR Pfam; PF01437; PSI; 1.
DR Pfam; PF01403; Sema; 1.
DR SMART; SM00423; PSI; 1.
KW Hypothetical protein.
SQ SEQUENCE 475 AA; 52298 MW; EE9090701AE2264 CRC64;

Query Match 50.0%; Score 46; DB 4; Length 475;
Best Local Similarity 61.5%; Pred. No. 22;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 YVOHQKMKPMIQ 13
Db 193 LYOPQATRPWQ 205

RESULT 15

O9C0B8 PRELIMINARY; PRT; 893 AA.
AC O9C0B8;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein KIAA1745 (Fragment).
GN KIAA1745.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21082932; PubMed=11214970;
RX Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Ref. 7:347-351(2000).
DR EMBL; AB051532; BAB21836.1; -
DR InterPro; IPR003659; Plexin-like.
DR InterPro; IPR002165; Plexin_repeat.
DR InterPro; IPR001627; Sema.
DR Pfam; PF01437; PSI; 1.
DR Pfam; PF01403; Sema; 1.
DR SMART; SM00423; PSI; 1.
DR SMART; SM00630; Sema; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 893 AA; 98361 MW; D6C6C48DE5524F14 CRC64;

Query Match 50.0%; Score 46; DB 4; Length 893;
Best Local Similarity 61.5%; Pred. No. 42;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 YVOHQKMKPMIQ 13
Db 611 LYOPQATRPWQ 623

Search completed: July 30, 2003, 16:29:20
Job time : 29.1429 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:10:10 ; Search time 5.2381 Seconds

(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-1
Sequence: 1 VYHOKAKKFWIQPKT 16

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query	Match Length	DB ID	Description
1	92	100.0	222	1	CAS2_BOVIN
2	77	83.7	223	1	CAS2_CAPRI
3	77	83.7	223	1	CAS2_SHEEP
4	56	60.9	235	1	CAS2_PIG
5	50	54.3	193	1	CAS2_CAMDR
6	46	50.0	180	1	CAS2_RABIT
7	46	50.0	832	1	SMAB_HUMAN
8	44	47.8	291	1	PCTU_MOUSE
9	44	47.8	389	1	CHS1_LYCES
10	44	47.8	389	1	CHS2_LYCES
11	44	47.8	389	1	CHS2_SOLTU
12	44	47.8	389	1	CHS2_PETHY
13	44	47.8	389	1	CHS2_SOLTU
14	44	47.8	389	1	CHS2_SOLTU
15	44	47.8	389	1	CHS2_IPOPU
16	44	47.8	389	1	CHS2_PETHY
17	44	47.8	389	1	CHS2_PETHY
18	44	47.8	390	1	CHS2_PETHY
19	44	47.8	391	1	CHS2_PETHY
20	44	47.8	400	1	CHS2_PETHY
21	44	47.8	400	1	CHS2_PETHY
22	44	47.8	401	1	CHS2_PETHY
23	44	47.8	401	1	CHS2_PETHY
24	44	47.8	401	1	CHS2_PETHY
25	44	47.8	401	1	CHS2_PETHY
26	44	47.8	401	1	CHS2_PETHY
27	44	47.8	401	1	CHS2_PETHY
28	44	47.8	401	1	CHS2_PETHY
29	44	47.8	401	1	CHS2_PETHY
30	44	47.8	401	1	CHS2_PETHY
31	44	47.8	401	1	CHS2_PETHY
32	44	47.8	401	1	CHS2_PETHY
33	44	47.8	401	1	CHS2_PETHY

ALIGNMENTS

34	42	45.7	186	1	ESM7_DROME	P13097_drosophila
35	42	45.7	224	1	ESM7_DROME	Q01068_drosophila
36	42	45.7	389	1	KSS5_BCOLI	P42218_escherichia
37	42	45.7	404	1	KIME_SCHPO	Q09780_schizosacch
38	42	45.7	589	1	ARSE_HUMAN	P51690_homo_sapien
39	41	44.6	195	1	VGIL_HSVNG	P52510_marek's dis
40	41	44.6	327	1	MOAA_SYNY3	O55369_synecocyst
41	41	44.6	345	1	TM04_MOUSE	Q91186_mus_musculu
42	41	44.6	392	1	CHS1_SECE	P53418_bacale cere
43	41	44.6	394	1	CHS2_SECE	P53415_bacale cere
44	41	44.6	398	1	CHS1_HORVU	P26018_hordeum vul
45	41	44.6	398	1	CHS1_ORYZA	P48405_oryza sativ

RESULT 1

AC	CAS2_BOVIN	STANDARD;	PRT;	222 AA.
DR	21-JUL-1986 (Rel. 01, Created)			
DT	01-JUL-1989 (Rel. 11, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Alpha-s2 casein precursor [contains: Casocidin-I].			
GN	CSN1S2.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OX	Bovidae; Bovinae; Bos.			
NCBI_TaxID=9913;				
RP	SEQUENCE FROM N.A.			
KA	MEDLINE=86186989; PubMed=2833669;			
RA	Stewart A.F., Bonning J., Beattie C.W., Shah F., Willis I.M.,			
RA	Mackinlay A.G.;			
RT	"Complete nucleotide sequences of bovine alpha S2- and beta-casein			
RT	CDNAs: comparisons with related sequences in other species.";			
RM	Mol. Biol. Evol. 4:231-241(1987).			
RN	[2]			
RP	SEQUENCE OF 16-222 (A ALLELE).			
RC	TISSUE=Milk;			
RX	MEDLINE=77185633; PubMed=862906;			
RA	Brignon G., Ribadeau-Dumas B., Mercier J.-C., Pellissier J.-P.,			
RA	Das B.C.;			
RT	"Complete amino acid sequence of bovine alphaS2-casein.";			
RT	FEBS Lett. 76:274-279(1977).			
RN	[3]			
RP	PARTIAL SEQUENCE (D ALLELE).			
RC	TISSUE=Milk;			
RX	MEDLINE=79239837; PubMed=469044;			
RA	Grosclaude F., Joudrier P., Mahe M.-F.;			
RT	"A genetic and biochemical analysis of a polymorphism of bovine alpha			
RT	S2-casein.";			
RT	J. Dairy Res. 46:211-213(1979).			
RN	[4]			
RP	SEQUENCE OF 165-203, AND CHARACTERIZATION OF CASOCIDIN.			
RC	TISSUE=Milk;			
RX	MEDLINE=9600204; PubMed=7556666;			
RA	Zucht H.-D., Raída M., Adernann K., Meagert H.-J., Forssmann W.-G.;			
RT	"Casocidin-I: a casein-alpha s2 derived peptide exhibits antibacterial			
RT	activity.";			
RT	FEBS Lett. 372:185-188(1995).			
CC	- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT			
CC	CALCIUM PHOSPHATE.			
CC	- FUNCTION: CASOCIDIN-I INHIBITS THE GROWTH OF ESCHERICHIA COLI AND			
CC	STAPHYLOCOCCUS CARNOBES.			
CC	- SUBCELLULAR LOCATION: Extracellular.			
CC	- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.			
CC	- MASS SPECTROMETRY: MW=4870; METHOD=Electrospray; RANGE=165-203.			
CC	- POLYMORPHISM: AT LEAST TWO ALLELES EXIST. THE SEQUENCE OF THE A			
CC	ALLELE IS SHOWN HERE. THE D ALLELE SEQUENCE DIFFERS FROM THAT			
CC	SHOWN IN HAVING A DELETION OF NINE RESIDUES, WHICH MAY BE 49-58,			

CC 50-59, OR 51-60.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC -1- DATABASE: NAME=Protein Spotlight;
 CC NOTE=Issue 16 of November 2001;
 CC WWW="http://www.expasy.org/spotlight/articles/sp16016.html".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M16644; AAA0479.1; -
 CC PIR; JQ2008; KABOS2.
 CC InterPro; IPR001588; Casein.
 CC Pfam; PF00363; caseins; 2.
 CC PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
 CC Milk; Phosphorylation; Signal; Repeat; Antibiotic.
 CC
 CC SIGNAL 1 15
 CC CHAIN 16 222 ALPHA-S2 CASEIN.
 CC PEPTIDE 165 203 CASOCIDIN-1.
 CC REPEAT 158 222
 CC MOD_RES 23 23 PHOSPHORYLATION.
 CC MOD_RES 24 24 PHOSPHORYLATION.
 CC MOD_RES 25 25 PHOSPHORYLATION.
 CC MOD_RES 31 31 PHOSPHORYLATION.
 CC MOD_RES 71 71 PHOSPHORYLATION.
 CC MOD_RES 72 72 PHOSPHORYLATION.
 CC MOD_RES 73 73 PHOSPHORYLATION.
 CC MOD_RES 76 76 PHOSPHORYLATION.
 CC MOD_RES 144 144 PHOSPHORYLATION.
 CC MOD_RES 146 146 PHOSPHORYLATION.
 CC CONFLICT 102 102 O -> E (IN REF. 2).
 CC SEQUENCE 222 AA; 26019 MW; 81E7408AF1C12FC CRC64;
 CC
 CC Query Match 100.0%; Score 92; DB 1; Length 222;
 CC Best Local Similarity 100.0%; Pred. No. 1e-07;
 CC Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC 1 VYOHOKAMKPMIOPKT 16
 CC 198 VYOHOKAMKPMIOPKT 213
 CC
 CC RESULT 2
 CC CAS2_CAPHI STANDARD; PRT; 223 AA.
 CC ID CAS2_CAPHI
 CC AC P33049;
 CC DT 01-OCT-1993 (Rel. 27, Created)
 CC DT 01-OCT-1993 (Rel. 27, Last sequence update)
 CC DE 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Alpha-S2 casein precursor (Alpha-S2-CN).
 CC GN CSNIS2.
 CC OS Capra hircus (Goat).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC OC Bovidae; Caprinae; Capra.
 CC NCBI_TaxID=9925;
 CC RN 1)
 CC RP SEQUENCE FROM N.A.
 CC RA Bouniol C.; Brignon G.; Mahe M.-F.; Prinz C.;
 CC "Characterization of goat alpha 2-caseins A and B: further
 CC evidence of the phosphorylation code of caseins";
 CC Protein Seq. Data Anal. 5:213-218 (1993).
 CC RL Gene 125:235-236 (1993).
 CC

RN [3]
 RP SEQUENCE FROM N.A. (VARIANT C).
 RX MEDLINE=9503056; PubMed=7943951;
 RA Bouniol C.; Brignon G.; Mahe M.-F.; Prinz C.;
 RT "Biochemical and genetic analysis of variant C of caprine alpha s2-
 RL casein (Capra hircus).";
 RL Ann. Genet. 25:173-177 (1994).
 CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
 CC CALCIUM PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
 CC -1- POLYMORPHISM: THREE ALLELES OF ALPHA-S2 CASEIN ARE KNOWN: A, B AND
 CC C. THE FREQUENCIES OF THE ALLELES IS ESTIMATED TO BE 0.85, 0.04
 CC AND 0.11 IN THE FRENCH DAIRY BREEDS 'ALPINE' AND 'SAANEY'.
 CC
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X65160; CAA46278.1; -
 CC EMBL; S74171; AAB32166.1; -
 CC PIR; I46995; I46995.
 CC PIR; S33881; JN0547.
 CC InterPro; IPR001588; Casein.
 CC Pfam; PF00363; caseins; 2.
 CC PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
 CC Milk; Phosphorylation; Signal; Repeat; Polymorphism.
 CC
 CC SIGNAL 1 15
 CC CHAIN 16 223 ALPHA-S2 CASEIN.
 CC REPEAT 77 141
 CC MOD_RES 159 223
 CC MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 77 77 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 145 145 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 147 147 PHOSPHORYLATION (BY SIMILARITY).
 CC VARIANT 79 79 E -> K (IN VARIANT B).
 CC VARIANT 182 182 K -> I (IN VARIANT C).
 CC SEQUENCE 223 AA; 26389 MW; 187DEF42FD688291 CRC64;
 CC
 CC Query Match 83.7%; Score 77; DB 1; Length 223;
 CC Best Local Similarity 87.5%; Pred. No. 2.6e-05;
 CC Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC 1 VYOHOKAMKPMIOPKT 16
 CC 199 VDOHOKAMKPMIOPKT 214
 CC
 CC RESULT 3
 CC CAS2_SHEEP STANDARD; PRT; 223 AA.
 CC ID CAS2_SHEEP
 CC AC P04654;
 CC DT 13-AUG-1987 (Rel. 05, Created)
 CC DT 13-AUG-1987 (Rel. 05, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Alpha-S2 casein precursor.
 CC GN CSNIS2.
 CC OS Ovis aries (Sheep).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC OC Bovidae; Caprinae; Ovis.
 CC NCBI_TaxID=9940;
 CC RN [1]
 CC RP


```

RP SEQUENCE FROM N.A.
RX MEDLINE=86104467; PubMed=1002499;
RA Boisenard M., Petrisseant G.;
RL "Complete sequence of ovine alpha s2-casein messenger RNA.";
RL Biochimie 67:1043-1051(1985).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCITRIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL, X03238; CA26983.1; -.
DR PIR, A25070; KASHS2.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN ALPHA BETA; 1.
KW Milk; Phosphorylation; Repeat; Signal.
FT SIGNAL 1 15
FT CHAIN 16 223 ALPHA-S2 CASEIN.
FT REPEAT 77 141
FT REPEAT 159 223
FT MOD_RES 23 23 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 24 24 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 25 25 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 72 72 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 73 73 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 74 74 PHOSPHORYLATION (POTENTIAL).
FT VARIANT 64 64 D -> N.
SQ SEQUENCE 223 AA; 26332 MW; 67212935E27426D7 CRC64;

Query Match 83.7%; Score 77; DB 1; Length 223;
Best Local Similarity 87.5%; Pred. No. 2.6e-05;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VYQHOKMKPMIOPKT 16
Db 199 VDHOKMKPMIOPKT 214

RESULT 4
CAS2_PIG STANDARD; PRT; 235 AA.
AC P39036;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-S2 casein precursor.
GN CSNIS2.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9623;
RN 11
RP SEQUENCE FROM N.A.
RX TISSUE=Mammary gland;
RX MEDLINE=92367960; PubMed=1503276;
RA Alexander L.J., Das Gupta N.A., Beattie C.W.;
RL "The sequence of porcine alpha s2-casein cDNA.";
RL Anim. Genet. 23:365-367(1992).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCITRIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
-----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X54975; CA38719.1; -.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN ALPHA BETA; 1.
KW Milk; Phosphorylation; Signal.
FT SIGNAL 1 15
FT CHAIN 16 235 ALPHA-S2 CASEIN.
FT MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 70 70 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 71 71 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 235 AA; 27570 MW; C903B760D184C14C CRC64;

Query Match 60.9%; Score 56; DB 1; Length 235;
Best Local Similarity 68.8%; Pred. No. 0.064;
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYQHOKMKPMIOPKT 16
Db 208 VDHOKMKPMIOPKT 223

RESULT 5
CAS2_CAMDR STANDARD; PRT; 193 AA.
AC O97944;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Alpha-S2 casein precursor.
GN CSNIS2.
OS Camelus dromedarius (Dromedary) (Arabian camel).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9638;
RN 11
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX STRAIN=Somali; TISSUE=udder;
RX MEDLINE=98291310; PubMed=9627840;
RA Kappeler S., Farah Z., Pahan Z.;
RL "Sequence analysis of Camelus dromedarius milk caseins.";
RL J. Dairy Res. 65:209-222(1998).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCITRIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AJ012629; CA410078.1; -.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN ALPHA BETA; FALSE_NEG.
-----

```

KM MILK: Phosphorylation; Signal.
 FT SIGNAL 1 15 ALPHA-S2 CASEIN.
 FT CHAIN 16 193
 SQ SEQUENCE 193 AA; 22964 MW; 2843356F8F2ED13 CRC64;
 Query Match 54.3%; Score 50; DB 1; Length 193;
 Best Local Similarity 72.7%; Pred. No. 0.48;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VYOHOKAMKPM 11
 DB 176 VYOHOKAMTWM 186

RESULT 6
 CAS2 RABIT STANDARD; PRT; 180 AA.
 AC P50418;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE Alpha-S2A casein precursor.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxId=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=New Zealand white; TISSUE=Mammary gland;
 RC MEDLINE=94107245; PubMed=8280077;
 RA Dawson S.P., Wilde C.J., Tighe P.J., Mayer R.J.;
 RT "Characterization of two novel casein transcripts in rabbit mammary gland.";
 RL Biochem. J. 296:777-784(1993).
 CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
 CC -1- CALCIUM PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: Mammary gland; MILK.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X76907; CAA54228.1; -
 DR PIR; S39775; S39775.
 DR InterPro; IPR001588; Casein.
 DR PROSITE; PS00306; CASEIN ALPHA_BETA, 1.
 KM MILK: Phosphorylation; Signal.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 16 180 ALPHA-S2A CASEIN.
 SQ SEQUENCE 180 AA; 21330 MW; 4CD1EBB9F9D1CF53 CRC64;
 Query Match 50.0%; Score 46; DB 1; Length 180;
 Best Local Similarity 56.2%; Pred. No. 2;
 Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 1 VYOHOKAMKPM 16
 DB 99 VYOHOKAMTWM 114

RESULT 7
 SMAB_HUMAN STANDARD; PRT; 832 AA.
 AC Q9N9R2; Q9N9P9; Q96FK5; Q9C0B8; Q9H691; Q9N9M6; Q9N9N0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Semaphorin 4B precursor.
 GN SEMA4B OR KIAA1745.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RC MEDLINE=21082932; PubMed=11214970;
 RX Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:347-355(2000).
 RN [2]
 RP SEQUENCE OF 160-832 FROM N.A., AND VARIANT ALA-792.
 RA Carin L., Estivill X., Escarceller M., Sumoy L.;
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBD databases.
 RN [3]
 RP SEQUENCE OF 358-832 FROM N.A.
 RA Kawabata A., Hiki T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
 RA Tanaka T., Nakamura Y., Isegaki T., Sugano S.;
 RA "NEO human cDNA sequencing project.";
 RT Submitted (AUG-2000) to the EMBL/Genbank/DBD databases.
 RL [4]
 RP SEQUENCE OF 303-832 FROM N.A., AND VARIANT ALA-792.
 RC TISSUE=Colon, and Pancreas;
 RC MEDLINE=22386257; PubMed=12477932;
 RX Strussberg R.L., Felngold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Usdin T.B., Tohyuki S., Carninci P., Frange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Halys S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Hellon E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
 CC SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
 CC -1- SIMILARITY: Contains 1 semadomin.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
 CC -1- CAUTION: Ref.3 sequence differs from that shown in position 709
 CC onward due to a frameshift.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB051532; BAB21836.1; -
 DR EMBL; AL339080; CAB98204.1; -
 DR EMBL; AL339081; CAB98205.1; -
 DR EMBL; AL339082; CAB98206.1; -
 DR EMBL; AK026133; BAB15372.1; ALT_FRAME.

DR EMBL; BC010701; AAH10701.1; ALT_INIT.
 DR Genew; HGNC:10730; SEMA4B.
 DR InterPro; IPR003659; Plexin-like.
 DR InterPro; IPR001651; Plexin_repeat.
 DR Pfam; PF01437; PSI; 1.
 DR Pfam; PF01403; Sema; 1.
 DR SMART; SM00423; PSI; 1.
 DR SMART; SM00630; Sema; 1.
 DR Transmembrane; immunoglobulin domain; Multigene family; Neurogenesis;
 DR Developmental protein; Glycoprotein; Signal; Polymorphism.
 FT SIGNAL 1
 FT CHAIN 39 832
 FT DOMAIN 39 712
 FT TRANSMEM 713 733
 FT DOMAIN 734 832
 FT DOMAIN 65 502
 FT DOMAIN 520 574
 FT DOMAIN 599 658
 FT DOMAIN 753 776
 FT DISULFID 606 651
 FT CARBOHYD 64 64
 FT CARBOHYD 91 91
 FT CARBOHYD 160 160
 FT CARBOHYD 405 405
 FT CARBOHYD 520 520
 FT CARBOHYD 625 625
 FT CARBOHYD 792 792
 FT VARIANT S -> A.
 FT CONFLICT 160 160
 FT CONFLICT 565 565
 FT SEQUENCE 832 AA; 92192 MW; 29D58C1DD5E1C6B CRC64;
 Query Match 50.0%; Score 46; DB 1; Length 832;
 Best Local Similarity 61.5%; Pred. No. 9;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VYQHKAKKPMIQ 13
 DB 550 LYQPLATRPWQ 562
 RESULT 8
 PCTL_MOUSE STANDARD; PRT; 291 AA.
 AC Q9JND3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE PCTP-like protein (PCTP-L) (Star-related lipid transfer protein 10)
 DE (STARL10) (STAR domain-containing protein 10) (serologically defined
 DE colon cancer antigen 28 homolog).
 DE STARD10 OR SDCAG28 OR SDCAG28 OR PCTPL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=20297013; PubMed=10819773;
 RA Yamahaka M., Koga M., Tanaka H., Nakamura Y., Ohta H., Yomogida K.,
 RA Tachikawa J., Iguchi N., Nojima H., Nozaki M., Matsumiya K.,
 RA Okuyama A., Toshimori K., Nishimune Y.;
 RT "Molecular cloning and characterization of phosphatidylcholine
 RT transfer protein-like protein gene expressed in murine haploid germ
 RT cells";
 RL Biol. Reprod. 62:1694-1701(2000).
 CC -1- FUNCTION: May play specific roles in sperm maturation or
 CC fertilization.
 CC -1- SUBCELLULAR LOCATION: In testis was predominantly detected at the
 CC flagella of elongated spermatids, with a strong signal also found

CC at the tail of epididymal sperm.
 CC -1- TISSUE SPECIFICITY: Testis, kidney, liver, and intestine with the
 CC highest level in the testis.
 CC -1- DEVELOPMENTAL STAGE: During male germ cell development, it was
 CC detected first in the 23-day-old mouse testis, and the signal
 CC increased with age.
 CC -1- SIMILARITY: Contains 1 STAR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB031550; BA92233.1; -
 DR MGP; MGI:1860093; Stard10.
 DR InterPro; IPR002913; STAR.
 DR Pfam; PF01852; STAR; 1.
 DR SMART; SM00234; STAR; 1.
 DR PROSITE; PS50848; STAR; 1.
 FT DOMAIN 14 224
 FT SEQUENCE 291 AA; 32951 MW; 598467C430D97080 CRC64;
 Query Match 47.8%; Score 44; DB 1; Length 291;
 Best Local Similarity 46.2%; Pred. No. 6;
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 3 QHOKAKKPMIQ 15
 DB 228 KHPKPMHP 240
 RESULT 9
 CHSL_LYCSES STANDARD; PRT; 389 AA.
 AC P23418;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUN-1999 (Rel. 38, Last annotation update)
 DE Chalcone synthase 1 (EC 2.3.1.74) (Naringenin-chalcone synthase 1).
 DE CHSL.
 GN Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Solanaceae; Solanum.
 OC NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cotyledon, Hypocotyl, and Leaf;
 RX MEDLINE=91117196; PubMed=1980524;
 RA O'Neill S.D., Tong Y., Spoerlein B., Forkmann G., Yoder J.I.;
 RT "Molecular genetic analysis of chalcone synthase in Lycopersicon
 RT esculentum and an anthocyanin-deficient mutant";
 RL Mol. Gen. Genet. 224:279-288(1990).
 CC -1- FUNCTION: The primary product of this enzyme is 4',2',4',6'-
 CC tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4 coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CHSA_PETHY
ID CHSA_PETHY STANDARD; PRT; 389 AA.
AC P08894;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase A (EC 2.3.1.74) (Naringenin-chalcone synthase A).
GN CHSA.
OS Petunia hybrida (Petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Petunia.
OX NCBI_TaxID=4102;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Violet 30;
RX MEDLINE=86286540; PubMed=3016642;
RA Koes R.E., Spelt C.E., Reif H.J., van den Elzen P.J.M., Velkamp E.,
Mol J.N.M.,
RT "Floral tissue of Petunia hybrida (V30) expresses only one member of
the chalcone synthase multigene family."
RL Nucleic Acids Res. 14:5229-5239(1986).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Violet 30; TISSUE=leaf;
RX MEDLINE=90034197; PubMed=2806915;
RA Koes R.E., Spelt C.E., van den Elzen P.J.M., Mol J.N.M.,
RT "Cloning and molecular characterization of the chalcone synthase
multigene family of Petunia hybrida."
RL Gene 81:245-257(1989).
CC -1- FUNCTION: The primary product of this enzyme is 4,2',4'',6''-
tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
which can under specific conditions spontaneously isomerize into
naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
OF WHICH ARE BRIGHTLY COLORED.
CC -1- TISSUE SPECIFICITY: MAJOR EXPRESSED MEMBER OF THE GENE FAMILY IN
VARIOUS FLORAL TISSUES AND IN SEEDLINGS TREATED WITH UV LIGHT. IT
IS RELATIVELY LOW EXPRESSED IN TISSUE CULTURE MATERIAL.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; X04080; CAA27718.1; -;
DR EMBL; X14591; CAA32731.1; -;
DR PIR; A23643; SYRPN.
DR PIR; J80308; SYRPN.
DR InterPro: IPR001099; N-C synthase.
DR Pfam: PF00195; Chal_stil_synth; 1.
DR Pfam: PF02797; Chal_stil_synth; 1.
DR ProDom: PD000453; N-C_synthase; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase;
KW Multigene family.
KW ACT SITE
FT ACT SITE 164
FT CONFLICT 128 V -> F (IN REF. 2).
SQ SEQUENCE 389 AA; 42525 MW; 0221B8DA1B56EAF6 CRC64;
Query Match 47.8%; Score 44; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 8.8;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
OY 5 OKAMKPMIOPKT 16
|||||
OY

Db 111 OKAIKEWGPERS 122
RESULT 13
ID CHSA_SOLUTU STANDARD; PRT; 389 AA.
AC 04136;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase 1A (EC 2.3.1.74) (Naringenin-chalcone synthase 1A).
GN CHS1A.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Red Pontiac;
RA Jeon J.-H., Kim H.-S., Choi K.-H., Young Y.-H., Young H., Byun S.-M.,
RT "Characterization of two members of the chalcone synthase gene family
from Solanum tuberosum L.";
RL (In) Plant Gene Register PGR96-027.
CC -1- FUNCTION: The primary product of this enzyme is 4,2',4'',6''-
tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
which can under specific conditions spontaneously isomerize into
naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; U47739; AAB67734.1; -;
DR InterPro: IPR001099; N-C synthase.
DR Pfam; PF00195; Chal_stil_synth; 1.
DR Pfam; PF02797; Chal_stil_synth; 1.
DR ProDom: PD000453; N-C_synthase; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase;
KW Multigene family.
KW ACT SITE
FT ACT SITE 164
FT CONFLICT 164
SQ SEQUENCE 389 AA; 42562 MW; B181D9C6B9170A34 CRC64;
Query Match 47.8%; Score 44; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 8.8;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
OY 5 OKAIKEWGPERS 122
|||||
OY
RESULT 14
ID CHSB_SOLUTU STANDARD; PRT; 389 AA.
AC 043163;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase 1B (EC 2.3.1.74) (Naringenin-chalcone synthase 1B).
GN CHS1B.
OS Solanum tuberosum (Potato).
OY

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Red Pontiac;
RA Jeon J.-H., Kim H.-S., Choi K.-H., Joung Y.-H., Joung H., Byun S.-M.;
RT "Characterization of two members of the chalcone synthase gene family
RT from Solanum tuberosum L."
RL (in) Plant Gene Register PGR96-027.
CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS. A LARGE CLASS OF SECONDARY PLANT METABOLITES. MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U47740; AAB67735.1; -
CC InterPro; IPR001099; N-C synthase.
CC Pfam; PF00195; Chal_still_synth; 1.
CC DR PIR; T07797; Chal_still_synth; 1.
CC PRODB: PP000453; N-C synthase; 1.
CC PROSITE; PS00441; CHALCONE_SYNTH; 1.
CC Flavonoid biosynthesis; Transferase; Acyltransferase;
CC MultiGene family.
CC ACT SITE 164
CC FT ACT SITE 164
CC SQ SEQUENCE 389 AA; 42548 MW; E7334A1A34B8D1CC CRC64;

Query Match 47.8%; Score 44; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 8.8;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAKKPMIOPKT 16
DB 111 OKAIKEWGQPKS 122

```

```

RL Plant Cell Physiol. 38:754-758(1997).
CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS. A LARGE CLASS OF SECONDARY PLANT METABOLITES. MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB001827; BAA21789.1; -
CC PIR; T07799; T07799.
CC InterPro; IPR001099; N-C synthase.
CC Pfam; PF00195; Chal_still_synth; 1.
CC DR PIR; T07797; Chal_still_synth; 1.
CC PRODB: PP000453; N-C synthase; 1.
CC PROSITE; PS00441; CHALCONE_SYNTH; 1.
CC Flavonoid biosynthesis; Transferase; Acyltransferase;
CC MultiGene family.
CC ACT SITE 164
CC FT ACT SITE 164
CC SQ SEQUENCE 389 AA; 42677 MW; 0CC81302CCA8A56A CRC64;

Query Match 47.8%; Score 44; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 8.8;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAKKPMIOPKT 16
DB 111 OKAIKEWGQPKS 122

```

Search completed: July 30, 2003, 16:24:28
Job time : 6.2381 secs

```

RESULT 15
CHSE_IPOPU
ID CHSE_IPOPU STANDARD; PRT; 389 AA.
AC 022047;
DT 15-JUL-1999 (rel. 38; Created)
DT 15-JUL-1999 (rel. 38; Last sequence update)
DT 28-FEB-2003 (rel. 41; Last annotation update)
DE Chalcone synthase E (EC 2.3.1.74) (Naringenin-chalcone synthase E)
DE (CHS-E)
GN CHSE
OS Ipomoea purpurea (Common morning glory).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Convolvulaceae; Ipomoea.
OX NCBI_TaxID=4121;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. FR-35; TISSUE=flower buds;
RX MEDLINE=97393496; PubMed=9249990;
RA Fukada-Tanaka S., Hoshino A., Hataomi Y., Habu Y., Hasebe M.,
RA Iida S.;
RT "Identification of new chalcone synthase genes for flower pigmentation
RT in the Japanese and common morning glories."

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:55 ; Search time 10.1905 Seconds
(without alignments)
150.994 Million cell updates/sec

Title: US-09-787-070-1

Perfect score: 92

Sequence: 1 VYHQKMKPWIOPKT 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	222	1	KABOS2
2	77	83.7	223	1	KASHS2
3	77	83.7	223	2	UN0547
4	77	83.7	223	2	146995
5	56	60.9	235	2	A48383
6	48	52.2	418	2	F72370
7	46.5	50.5	363	2	T23156
8	46.5	50.5	364	2	T29153
9	46	50.0	180	2	B58930
10	46	50.0	265	2	B58930
11	45	48.9	311	2	T15997
12	44	47.8	329	2	S18136
13	44	47.8	318	2	H75008
14	44	47.8	340	2	T10957
15	44	47.8	389	1	SYRUCJ
16	44	47.8	389	1	SYRUCJ
17	44	47.8	389	1	SYRUCJ
18	44	47.8	389	1	SYRUCJ
19	44	47.8	389	1	SYRUCJ
20	44	47.8	389	1	SYRUCJ
21	44	47.8	389	1	SYRUCJ
22	44	47.8	389	1	SYRUCJ
23	44	47.8	389	1	SYRUCJ
24	44	47.8	389	1	SYRUCJ
25	44	47.8	389	1	SYRUCJ
26	44	47.8	389	1	SYRUCJ
27	44	47.8	389	1	SYRUCJ
28	44	47.8	389	1	SYRUCJ
29	44	47.8	389	1	SYRUCJ

30	42.5	46.2	590	1	CYPRA1
31	42	45.7	143	1	KEMS
32	42	45.7	166	2	S03626
33	42	45.7	217	2	T04292
34	42	45.7	224	2	D46177
35	42	45.7	248	2	T01880
36	42	45.7	327	2	C71146
37	42	45.7	389	1	S36553
38	42	45.7	404	1	S62440
39	42	45.7	589	2	T37187
40	42	45.7	1091	2	T34107
41	42	45.7	232	2	B97147
42	41	44.6	324	2	S58061
43	41	44.6	327	1	S75716
44	41	44.6	367	2	T52363
45	41	44.6	398	2	S16275

ALIGNMENTS

RESULT 1
KABOS2
alpha-s2-casein precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 01-Sep-1981 #sequence revision 03-Feb-1994 #text_change 22-Jun-1999
C:Accession: J02008; A25087; A91438; S66626; A03107
R:Groenen, M.A.M.; Dijkhof, R.J.M.; Verstege, A.J.M.; van der Poel, J.J.
Gene 123, 187-193, 1993
A:Title: The complete sequence of the gene encoding bovine alpha-s2-casein.
A:Reference number: J02008; MIMD:93154583; PMID:842858
A:Accession: J02008
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-222 <GRO>
A:Cross-References: GB:M94327
R:Stewart, A.F.; Bomling, J.; Beattie, C.W.; Shah, F.; Willis, I.M.; Mackinlay, A.G.
Mol. Biol. Evol. 4, 231-241, 1987
A:Title: Complete nucleotide sequences of bovine alpha-s2- and beta-casein cDNAs: comparison of the two genes.
A:Reference number: A93062; MIMD:88188989; PMID:2833669
A:Accession: A29087
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-222 <STE>
A:Cross-References: GB:M16644; MIMD:9162928; PIR:MAA30479.1; PID:9162929
R:Brignon, G.; Ribadeau Dumas, B.; Mercier, J.C.; Pelissier, J.P.; Das, B.C.
FEBS Lett. 76, 274-279, 1977
A:Title: Complete amino acid sequence of bovine alpha-s2-casein.
A:Reference number: A91438; MIMD:77185633; PMID:862906
A:Accession: A91438
A:Molecule type: protein
A:Residues: 16-101, 'EE', 104-222

A>Note: four fractions, previously designated s2, s3, s4, and s6, appear to have the same
these
R:Groenen, M.A.M.; Dijkhof, R.J.M.; Verstege, A.J.M.; van der Poel, J.J.
J. Dairy Res. 46, 211-213, 1979
A:Title: A genetic and biochemical analysis of a polymorphism of bovine alpha-s2-casein
A:Reference number: A92771; MIMD:9239837; PMID:469044
A:Accession: A92771
A:Contents: annotation; D allele
A>Note: The sequence of the D allele has a deletion of nine residues, which may be 49-58
R:Zucht, H.D.; Raada, M.; Adermann, K.; Maegert, H.J.; Forsmann, W.G.
FEBS Lett. 372, 185-188, 1995
A:Title: Casein alpha-s2 derived peptide exhibits antibacterial activity
A:Reference number: S66626; MIMD:96000204; PMID:755666
A:Accession: S66626
A:Molecule type: protein
A:Residues: 165-203 <ZUC>
C:Comment: The sequence of the A allele is shown.
C:Genetics:
A:Gene: alpha-s2
A:Map position: 6
A:Introns: 17/3; 26/3; 33/3; 47/3; 56/3; 65/3; 74/3; 82/3; 97/3; 138/3; 147/3; 156/3; 161/3

C:Superfamily: alpha-s2-casein

C:Keywords: mammary gland; milk; phosphoprotein

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-222/Product: alpha-s2-casein #status experimental <MAT>

F:23,24,25,31,71,72,73,76,144,146,158/Binding site: phosphate (Ser) (covalent) #status P

Query Match 100.0%; Score 92; DB 1; Length 222;

Best Local Similarity 100.0%; Pred. No. 1.7e-07;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYQHOKAKMPWIOPT 16
|||||

Db 199 VDHOKAKMPWIOPT 213

RESULT 2

KASHS2

alpha-s2-casein precursor - sheep

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 22-Jun-1999

C:Accession: A25070; S17856

R:Bouinard, M.; Petrisant, G.

Biochimie 67, 1043-1051, 1985

A:Title: Complete sequence of ovine alpha-s2-casein messenger RNA.

A:Reference number: A25070; MUID:86104467; PMID:3002499

A:Accession: A25070

A:Molecule type: mRNA

A:Residues: 1-223 <BO1>

A:Cross-References: GB:X03228; NID:G1238; PIDN:CAA26983.1; PID:G732894

A:Note: 64-Aen was also found

R:Bouinard, M.; Hue, D.; Bouinard, C.; Mercier, J.C.; Gaye, P.

Eur. J. Biochem. 201, 633-641, 1991

A:Title: Multiple mRNA species code for two non-allelic forms of ovine alpha-s2-casein.

A:Reference number: S17856; MUID:92037619; PMID:1935959

A:Accession: S17856

A:Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-93, 'T', '95-223 <BO2>

C:Superfamily: alpha-s2-casein

C:Keywords: mammary gland; milk; phosphoprotein

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-223/Product: alpha-s2-casein #status predicted <KAS>

F:23,24,25,32,55,72,73,74,77,145,147,159/Binding site: phosphate (Thr) (covalent) #status predicted

F:53,68,146,154,170,198/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 83.7%; Score 77; DB 1; Length 223;

Best Local Similarity 87.5%; Pred. No. 4.6e-05;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VYQHOKAKMPWIOPT 16
|||||

Db 199 VDHOKAKMPWIOPT 214

RESULT 3

alpha-s2-casein precursor - goat

C:Species: Capra aegagrus hircus (domestic goat)

C:Date: 31-Dec-1993 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: S33881; S33880; UN0547; S20620

R:Bouinard, C.; Brignon, G.; Mahe, M.F.; Printz, C.

Protein Seq. Data Anal. 5, 213-218, 1993

A:Title: Characterization of goat allelic alpha-s2-caseins A and B: further evidence of

A:Reference number: S33880

A:Accession: S33881

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-223 <BOU>

A:Note: allele B

A:Accession: S33880

A:Molecule type: protein

A:Residues: 16-78, 'E', '80-223 <BO2>

A:Note: sequence deduced from compositional analysis of peptides

R:Bouinard, C.

Gene 125, 235-236, 1993

A:Title: Sequence of the goat alpha-s2-casein-encoding cDNA.

A:Reference number: UN0547; MUID:93216130; PMID:8462880

A:Accession: UN0547

A:Molecule type: mRNA

A:Residues: 1-78, 'E', '80-223 <BO3>

A:Cross-References: EMBL:X65160; NID:G955; PIDN:CAA46278.1; PID:G956

A:Note: allele A

C:Superfamily: alpha-s2-casein

C:Keywords: mammary gland; milk; phosphoprotein

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-223/Product: alpha-s2-casein #status experimental <MAT>

F:23,24,25,55,72,73,74,77,145,147,159/Binding site: phosphate (Ser) (covalent) #status predicted

F:53,146/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 83.7%; Score 77; DB 2; Length 223;

Best Local Similarity 87.5%; Pred. No. 4.6e-05;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VYQHOKAKMPWIOPT 16
|||||

Db 199 VDHOKAKMPWIOPT 214

RESULT 4

alpha s2-casein C - goat

C:Species: Capra aegagrus hircus (domestic goat)

C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 13-Aug-1999

C:Accession: I46995

R:Bouinard, C.; Brignon, G.; Mahe, M.F.; Printz, C.

Antim. Genet. 25, 173-177, 1994

A:Title: Biochemical and genetic analysis of variant C of caprine alpha s2-casein (Capra

A:Reference number: I46995; MUID:95030556; PMID:7943951

A:Accession: I46995

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-223 <BOU>

A:Cross-References: GB:S74171; NID:G707033; PIDN:AB832166.1; PID:G707034

C:Superfamily: alpha-s2-casein

Query Match 83.7%; Score 77; DB 2; Length 223;

Best Local Similarity 87.5%; Pred. No. 4.6e-05;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VYQHOKAKMPWIOPT 16
|||||

Db 199 VDHOKAKMPWIOPT 214

RESULT 5

alpha s2-casein - pig

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 03-May-1996

C:Accession: A48383

R:Alexander, L.J.; Das Gupta, N.A.; Beattie, C.W.

Antim. Genet. 23, 365-367, 1992

A:Title: The sequence of porcine alpha s2-casein cDNA.

A:Reference number: A48383; MUID:92367960; PMID:1503276

A:Accession: A48383

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-235 <ALB>

A:Experimental source: mammary gland

A:Note: sequence inconsistent with the nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBIN:110884; NCBI:P:110885)

C:Superfamily: alpha-s2-casein

Query Match 60.9%; Score 56; DB 2; Length 235;

Best Local Similarity 68.8%; Pred. No. 0.12;

Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;


```

Query Match      50.0%; Score 46; DB 2; Length 265;
Best Local Similarity 46.2%; Pred. No. 5.8;
Matches      6; Conservative      4; Mismatches      3; Indels      0; Gaps      0;
QY      2 YOHKAKMKPIOP 14
          |:::|::|
Db      134 YTOHKSIPWIDP 146

```

```

RESULT 11
T15997
hypothetical protein F09E5.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15997
R:Chisaoe, S.
submitted to the EMBL Data Library, September 1995
A:Description: The sequence of C. elegans cosmid F09E5.
A:Reference number: Z18444
A:Accession: T15997
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-311 <CHI>
A:Cross-references: EMBL:U07429, NID:g1019949, PID:g1019956, PIDN:AAA79347.1; CESP:F09E5
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F09E5.7
A:Introns: 22/3; 77/1; 238/2

OY      3 OHOKAMKPMTPKPT 16
|||::||| | | | |
DB      21 QHESMSKMWNGPPT 34

Query Match      48.9%; Score 45; DB 2; Length 311;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```

```

RESULT 12
S18136
maritangenin-chalcone synthase (EC 2.3.1.74) - garden petunia
C:Species: Petunia x hybrida (garden petunia)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-May-2000
C:Accession: S18136
R:van Tunen, A.J.
submitted to the EMBL Data Library, March 1989
A:Reference number: S18136
A:Accession: S18136
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-319 <1UN>
A:Cross-references: GDB:G14599; NID:G20537; PIDN:CAA32739.1; PID:G20538
C:Superfamily: chalcone synthase
C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match      47.8%; Score 44; DB 2; Length 319;
Best Local Similarity 66.7%; Pred. No. 15;
Matches      8; Conservative      2; Mismatches      2; Indels      0; Gaps      0;

QY      5 OKAMKPIQPKT 16
      |||||
DB      41 OKAIKMGQPKS 52

RESULT 13
H75008
hychochemical protein PAB1055 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 26-Jul-2000
C:Accession: H75008
R:anonymous, Genomecope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

```

A:Reference number: A75001
A:Accession: H75008
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1328 <RAM>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:q5458960; PIDN:CAB50510.1; PID:q545902
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1055
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB1055

Query Match 47.8%; Score 44; DB 2; Length 328;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VYQHQRKMKPWI 12
:::|::|
Db 109 IYRKRAIKPWI 120

RESULT 14
T10957
naringenin-chalcone synthase (EC 2.3.1.74) CHS-FL1 - common morning-glory
C:Species: Ipomoea purpurea (common morning-glory)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #ext_change 05-May-2000
C:Accession: T10957
R:Rauscher, M.D.; Tiffin, P.L.; Miller, R.E.
submitted to the EMBL Data Library January 1997
A:Description: Regulation of anthocyanin gene expression in Ipomoea purpurea.
A:Reference number: Z17229
A:Accession: T10957
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1340 <RAU>
A:Cross-references: EMBL:U74082; NID:G1658119; PID:G1658120
A:Experimental source: flower bud; WMAA genotype (dark flowered)
C:Genetics:
A:Gene: CHS-FL1
C:Superfamily: chalcone synthase
C:Keywords: acyltransferase; coenzyme A

Query Match 47.8%; Score 44; DB 2; Length 340;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 QRKMKPWIOTPKT 16
|||::|::|
Db 111 QRKMKWGQPKS 122

RESULT 15
SYBJCJ
naringenin-chalcone synthase (EC 2.3.1.74) J - garden petunia
N:Alternate names: chalcone synthase
C:Species: Petunia x hybrida (garden petunia)
C:Date: 30-Sep-1991 #sequence_revision 17-Mar-2000 #ext_change 05-May-2000
C:Accession: D72821; JS0309
R:Koes, R.E.; Spelt, C.E.; van den Elzen, P.J.M.; Mol, J.N.M.
Gene 81, 245-257, 1989
A>Title: Cloning and molecular characterization of the chalcone synthase multigene family
A:Reference number: JS0308; MUID:90034197; PMID:2806915
A:Accession: D72821
A:Molecule type: DNA
A:Residues: 1389 <KOE>
A:Cross-references: EMBL:X14597; NID:G20535; PIDN:CAA32737.1; PID:G20536
A:Experimental source: strain Violet 30, leaf
A:Molecule type: DNA
A:Residues: 150, 'D', 52-74, 'V', 76-228, 'I', 230-297, 'L', 299-389 <KOE>
A:Note: The sequence is revised in GenBank entry PHGSHJ release 114, (PIDN:CAA32737.1)
C:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavo-
C:Genetics:
A:Gene: chs

RESULT 15
SYBJCJ
naringenin-chalcone synthase (EC 2.3.1.74) J - garden petunia
N:Alternate names: chalcone synthase
C:Species: Petunia x hybrida (garden petunia)
C:Date: 30-Sep-1991 #sequence_revision 17-Mar-2000 #ext_change 05-May-2000
C:Accession: D72821; J50309
R:Koes, R.E.; Spelt, C.E.; van den Elzen, P.J.M.; Mol, J.N.M.
Gene 81, 245-257, 1989
A>Title: Cloning and molecular characterization of the chalcone synthase multigene family
A:Reference number: J50308; MUID:90034197; PMID:2806915
A:Accession: D72821
M:Molecule type: DNA
A:Residues: 1-389 <KOE>
A:Cross-references: EMBL:X14597; NID:G20535; PIDN:CAA32737.1; PID:G20536
A:Experimental source: strain Violet 30, leaf
A:Accession: J50309
M:Molecule type: DNA
A:Residues: 1-50, 'D', '52-74, 'V', '76-228, 'I', '230-297, 'L', '299-389 <KOE>
A:Note: the sequence is revised in GenBank entry BHCSJ release 114, (PIDN:CAA32737.1)
C:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavo-
C:Genetics:
A:Gene: chsJ

A: Map position: V
A: Note: chet is expressed in various floral tissues and UV illuminated seedlings
C: Superfamily: chalcone synthase
C: Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 47.8%; Score 44; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKMKPMTQPKT 16
|||:|||||:
Db 111 OKAIKMGQPKS 122

Search completed: July 30, 2003, 16:31:14
Job time : 11.1905 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:09:50 ; Search time 31.5238 Seconds
(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-1

Perfect score: 92

Sequence: 1 VYGHQKMKPWIOPKT 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	16	21	AAV88337
2	92	100.0	21	21	AAV88338
3	92	100.0	24	15	AAV88339
4	92	100.0	24	21	AAV88340
5	92	100.0	25	13	AAV88341
6	92	100.0	25	21	AAV88342
7	92	100.0	31	18	AAV88343
8	92	100.0	31	23	AAV88344
9	92	100.0	222	18	AAV88345

10	92	100.0	222	23	AAV88346	Bovine alpha-S2 ca
11	81.5	88.6	26	20	AAV88347	Bifidobacterium bi
12	77	83.7	223	23	AAV88348	Alpha-S2 casein pr
13	77	83.7	223	23	AAV88349	Coat alpha-S2 case
14	77	83.7	223	23	AAV88350	Capra hircus alpha
15	77	83.7	223	23	AAV88351	Sheep alpha-S2 cas
16	77	83.7	223	23	AAV88352	Alpha-S2 casein pr
17	58	63.0	19	18	AAV88353	Alpha-S2 casein pr
18	58	63.0	19	18	AAV88354	Bovine alpha-S2 ca
19	56	60.9	224	23	AAV88355	Pig alpha-S2 casei
20	53	57.6	9	17	AAV88356	Milk derived anti-
21	49	53.3	8	17	AAV88357	Milk derived anti-
22	48	52.2	103	22	AAV88358	Human immune/haema
23	48	52.2	191	21	AAV88359	Human secreted pro
24	48	52.2	328	22	AAV88360	S. epidermidis ope
25	48	52.2	332	23	AAV88361	Staphylococcus epi
26	48	52.2	575	22	AAV88362	Human EXMAD-17 SEQ
27	48	52.2	704	22	AAV88363	Human protein sequ
28	48	52.2	704	22	AAV88364	Human polypeptide
29	46	50.0	315	22	AAV88365	Novel human extrac
30	46	50.0	372	23	AAV88366	Human NS protein s
31	46	50.0	675	23	AAV88367	Human protein SEQ
32	46	50.0	681	21	AAV88368	Neuron-associated
33	46	50.0	791	23	AAV88369	Human NOV7 protein
34	46	50.0	832	22	AAV88370	Human gene 1 encod
35	46	50.0	832	23	AAV88371	Human albumin fusi
36	46	50.0	837	21	AAV88372	Human PRO1480 (UNQ
37	46	50.0	837	22	AAV88373	Human PRO1480 (UNQ
38	46	50.0	837	22	AAV88374	Human PRO1480 (UNQ
39	46	50.0	837	23	AAV88375	Protein of the inv
40	46	50.0	837	24	AAV88376	Hypoxia-induced pr
41	46	50.0	837	24	AAV88377	Human bladder can
42	46	50.0	837	24	AAV88378	Human secreted/tra
43	46	50.0	837	24	AAV88379	Novel human secret
44	46	50.0	837	24	AAV88380	Human secreted/tra
45	46	50.0	837	24	AAV88381	Human PRO1480 (UNQ

ALIGNMENTS

RESULT 1	AAV88337	standard; peptide; 16 AA.
ID	AAV88337	
XX	AAV88337	
AC	AAV88337	
XX	AAV88337	
DT	14-JUN-2000	(first entry)
XX	14-JUN-2000	
DB	Internal peptide fragment of bovine alpha-S2 casein #1.	
XX	Alpha-S2 casein; peptide production; biological fluid; milk; whey; blood;	
KW	antibacterial peptide; lactoferrin; antiviral; antitumour activity.	
XX		
OS	Bos sp.	
XX		
PN	WO2000015655-A1.	
XX		
PD	23-MAR-2000.	
XX		
PF	15-SEP-1999;	99WO-EP07002.
XX		
PR	15-SEP-1998;	98EP-0203107.
XX		
PR	08-JUN-1999;	99EP-0201815.
XX		
PA	(NIZO-) NIZO FOOD RES.	
XX		
PI	Visser S, Recio I;	
XX		
DR	WPI; 2000-271377/23.	
XX		
PT	Novel process for producing peptides with e.g. antimicrobial activity	
XX	from biological fluids such as milk, whey or blood comprises contacting	

PT Fluid with chromatographic medium to adsorb peptide domain of interest
XX
PS Claim 14, Page 22; 41pp; English.
XX
CC This sequence represents an internal fragment of bovine alpha-S2 caesin
CC protein. The peptide is an example of a peptide with antibacterial
CC activity that can be produced by the process of the invention. The
CC invention relates to a process for producing peptides from biological
CC fluids. The process comprises chromatography of the biological fluid, in
CC situ hydrolysis of selectively bound peptides, washing to remove unbound
CC peptide, and elution of the peptides of interest. The process is used for
CC producing peptides from biological fluids, such as milk, whey or blood.
CC For example, the process can be used to produce antibacterial peptides
CC derived from lactoferrin, using cheese whey as a starting material. The
CC peptides obtained have preferably antimicrobial and/or antiviral and/or
CC antitumour activity. The process of the invention is relatively simple
CC and generally economically and technically more attractive than those
CC methods previously used. The method provides high yield peptides with a
CC selected activity of interest without the need for intermediate
CC purification of the precursor protein.
SQ Sequence 16 AA;
Query Match 100.0%; Score 92; DB 21; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VYOHQKAMKPMIOPKT 16
DB 1 VYOHQKAMKPMIOPKT 16
RESULT 2
AA88338
ID AA88338 standard; peptide; 21 AA.
AC AAY88338;
XX
DT 14-JUL-2000 (first entry)
XX
DB Internal peptide fragment of bovine alpha-S2 caesin #2.
XX
KM Alpha-S2 caesin; peptide production; biological fluid; milk; whey; blood;
KM antibacterial peptide; lactoferrin; antiviral; antitumour activity.
XX
OS Bos sp.
XX
PN WO200015655-A1.
XX
PD 23-MAR-2000.
XX
PF 15-SEP-1999; 99WO-EP07002.
XX
PR 15-SEP-1998; 98EP-0203107.
XX
PR 08-JUN-1999; 99EP-0201815.
XX
PA (NIZO-) NIZO FOOD RES.
XX
PI Visser S, Recio I;
XX
DR WPI; 2000-271377/23.
XX
PT Novel process for producing peptides with e.g. antimicrobial activity
PT from biological fluids such as milk, whey or blood comprises contacting
PT fluid with chromatographic medium to adsorb peptide domain of interest
PT
PS Claim 14; Page 22; 41pp; English.
XX
CC This sequence represents an internal fragment of bovine alpha-S2 caesin
CC protein. The peptide is an example of a peptide with antibacterial
CC activity that can be produced by the process of the invention. The

CC invention relates to a process for producing peptides from biological
CC fluids. The process comprises chromatography of the biological fluid, in
CC situ hydrolysis of selectively bound peptides, washing to remove unbound
CC peptide, and elution of the peptides of interest. The process is used for
CC producing peptides from biological fluids, such as milk, whey or blood.
CC For example, the process can be used to produce antibacterial peptides
CC derived from lactoferrin, using cheese whey as a starting material. The
CC peptides obtained have preferably antimicrobial and/or antiviral and/or
CC antitumour activity. The process of the invention is relatively simple
CC and generally economically and technically more attractive than those
CC methods previously used. The method provides high yield peptides with a
CC selected activity of interest without the need for intermediate
CC purification of the precursor protein.
SQ Sequence 21 AA;
Query Match 100.0%; Score 92; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VYOHQKAMKPMIOPKT 16
DB 1 VYOHQKAMKPMIOPKT 16
RESULT 3
AAR60481
ID AAR60481 standard; peptide; 24 AA.
AC AAR60481;
XX
DT 19-APR-1995 (first entry)
XX
DE Peptide used in tranquilliser.
XX
KM Peptide; tranquilliser; food; foodstuff; mental stabilisation.
XX
OS Synthetic.
XX
PN JP06211689-A.
XX
PD 02-AUG-1994.
XX
PF 19-JAN-1993; 93JP-0024811.
XX
PR 19-JAN-1993; 93JP-0024811.
XX
PA (KANE) KANEBO LTD.
XX
DR WPI; 1994-283276/35.
XX
PT Synthetic peptide(s) used in tranquilliser - also used in the
PT production of a food for mental stabilisation.
XX
PS Claim 3; Page 1; 5pp; Japanese.
XX
CC The peptide may be used as a component of a tranquilliser which may
CC be administered orally and is low in toxicity. The peptide may also
CC be used as a component of a foodstuff which may be used for mental
CC stabilisation. See also AAR60480.
XX
SQ Sequence 24 AA;
Query Match 100.0%; Score 92; DB 15; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VYOHQKAMKPMIOPKT 16
DB 1 VYOHQKAMKPMIOPKT 16
RESULT 4

AAV88339
 ID AAV88339 standard; peptide; 24 AA.
 AC AAV88339;
 XX
 XX 14-JUN-2000 (first entry)
 XX
 XX Internal peptide fragment of bovine alpha-S2 caesin #2.
 XX
 XX Alpha-S2 caesin; peptide production; biological fluid; milk; whey; blood;
 KW antibacterial peptide; lactoferrin; antiviral; antitumour activity.
 XX
 OS Bos sp.
 XX
 XX WO200015655-A1.
 XX
 XX 23-MAR-2000.
 XX
 XX 15-SEP-1999; 99WO-EP07002.
 XX
 XX 15-SEP-1998; 98EP-0203107.
 PR 08-JUN-1999; 99EP-0201815.
 XX
 XX (NIZO-) NIZO FOOD RES.
 PA
 XX
 PI Visser S, Recio I;
 XX
 DR WPI; 2000-271377/23.
 XX
 PT Novel process for producing peptides with e.g. antimicrobial activity
 PT from biological fluids such as milk, whey or blood comprises contacting
 PT fluid with chromatographic medium to adsorb peptide domain of interest
 PT
 XX
 XX
 PS Claim 14; Page 22; 41pp; English.
 XX
 CC This sequence represents an internal fragment of bovine alpha-S2 caesin
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.
 CC
 XX
 XX Sequence 24 AA;
 SQ

Query Match 100.0%; Score 92; DB 21; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1,9e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKMKRPMIOPKT 16
 |||||
 DB 1 VYOHQKMKRPMIOPKT 16
 |||||

RESULT 5
 AAR20055
 ID AAR20055 standard; peptide; 25 AA.
 XX
 XX AAR20055;
 AC
 XX
 XX 25-MAR-2003 (updated)
 DT 26-MAR-1992 (first entry)
 XX

DE Casein peptide.
 XX
 XX Casein; platelet; aggregation; thrombosis; collagen.
 KW
 XX Synthetic.
 OS
 XX JP03255095-A.
 XX
 XX 13-NOV-1991.
 PD
 XX 02-MAR-1990; 90UP-0052554.
 PF
 XX 02-MAR-1990; 90UP-0052554.
 PR
 XX 02-MAR-1990; 90UP-0052554.
 XX
 XX (KANE) KANEBO LTD.
 PA
 XX
 XX WPI; 1992-002669/01.
 DR
 XX
 XX Casein peptide(s) for treating thrombosis - as inhibitors of
 PT platelet aggregation caused by adenosine-5'-diphosphate and
 PT collagen and as biochemical reagents
 PT
 XX
 XX Claim 1; Page 1; 8pp; Japanese.
 PS
 CC Ie25 may be omitted. The peptide and its salts have inhibiting
 CC activity against platelet aggregation caused by adenosine-5'-
 CC diphosphate and by collagen. They are useful for prevention and
 CC treatment of thrombosis and are also useful as a biochemical reagent.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 CC
 XX
 XX Sequence 25 AA;
 SQ

Query Match 100.0%; Score 92; DB 13; Length 25;
 Best Local Similarity 100.0%; Pred. No. 2e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKMKRPMIOPKT 16
 |||||
 DB 1 VYOHQKMKRPMIOPKT 16
 |||||

RESULT 6
 AAV88340
 ID AAV88340 standard; peptide; 25 AA.
 XX
 XX AAV88340;
 AC
 XX
 XX 14-JUN-2000 (first entry)
 DT
 DE C-terminal peptide fragment of bovine alpha-S2 caesin.
 XX
 XX Alpha-S2 caesin; peptide production; biological fluid; milk; whey; blood;
 KW antibacterial peptide; lactoferrin; antiviral; antitumour activity.
 KW
 XX
 OS Bos sp.
 XX
 XX WO200015655-A1.
 XX
 XX 23-MAR-2000.
 PD
 XX 15-SEP-1999; 99WO-EP07002.
 PF
 XX 15-SEP-1998; 98EP-0203107.
 PR 08-JUN-1999; 99EP-0201815.
 XX
 XX (NIZO-) NIZO FOOD RES.
 PA
 XX
 XX Visser S, Recio I;
 PI
 XX
 XX WPI; 2000-271377/23.
 DR
 PT Novel process for producing peptides with e.g. antimicrobial activity
 PT from biological fluids such as milk, whey or blood comprises contacting
 PT

PT fluid with chromatographic medium to adsorb peptide domain of interest
 XX
 PS Claim 14; Page 22; 41pp; English.
 CC This sequence represents a C-terminal fragment of bovine alpha-S2 casein
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumor activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.
 CC
 XX Sequence 25 AA;
 SQ
 Query Match 100.0%; Score 92; DB 21; Length 25;
 Best Local Similarity 100.0%; Pred. No. 28-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKPMIOPKT 16
 |||||
 1 VYQHOKAMKPMIOPKT 16
 Db
 RESULT 7
 AAW32219
 ID AAW32219 standard; peptide; 31 AA.
 XX
 AC AAW32219;
 DT 03-FEB-1998 (first entry)
 XX
 DE Alpha-S2 casein precursor C-terminal peptide fragment #5.
 XX
 KW Alpha-S2 casein precursor; growth promoting; mitogenic assay;
 KW platelet-derived growth factor; insulin-like derived growth factor;
 XX
 OS Synthetic.
 OS Bos taurus.
 XX
 PN MO9716460-A1.
 PD 09-MAY-1997.
 XX
 PF 31-OCT-1996; 96WO-GB02658.
 XX
 PR 31-OCT-1995; 95GB-0022302.
 XX
 PA (UWLI-) UNITV LIVERPOOL.
 PI Liu Q, Smith JA, Wilkinson MC;
 XX
 DR WPI; 1997-272048/24.
 XX
 PT Manufacture of medicament or foodstuff for promoting growth - using
 PT peptide(s) with a sequence identical to the C-terminal end of an
 PT alpha-S2 casein precursor
 XX
 PS Claim 12; Page 21; 33pp; English.
 XX
 CC The present sequence, which is substantially identical to the C-terminal
 CC end of an alpha-S2 casein precursor, was found after storage in PBS to
 CC exhibit growth promoting activity for rat mammary fibroblast cell line
 CC (Rame 27), which is not significantly stimulated by platelet-derived

CC growth factor or insulin-like growth factor. The activity of the peptide
 CC increased when maintained at alkaline pH. By way of contrast, alpha-S2
 CC casein was inactive in a mitogenic assay. This peptide may be used in the
 CC manufacture of a medicament or foodstuff for promoting growth in
 CC humans or animals.
 CC
 XX Sequence 31 AA;
 SQ
 Query Match 100.0%; Score 92; DB 18; Length 31;
 Best Local Similarity 100.0%; Pred. No. 2.58-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKPMIOPKT 16
 |||||
 7 VYQHOKAMKPMIOPKT 22
 Db
 RESULT 8
 AAB17466
 ID AAB17466 standard; peptide; 31 AA.
 XX
 AC AAB17466;
 DT 22-APR-2002 (first entry)
 XX
 DE Bovine alpha-S2 casein precursor protein C-terminal fragment #7.
 XX
 KW Bovine; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KW chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Bos sp.
 PN WO200202133-A2.
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Claim 9; Page 22; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is a peptide
 CC derived from the C-terminal of bovine alpha-S2 casein precursor protein.
 CC
 XX Sequence 31 AA;
 SQ
 Query Match 100.0%; Score 92; DB 23; Length 31;
 Best Local Similarity 100.0%; Pred. No. 2.58-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKPMIOPKT 16


```

DB      7 VYOHOKAMKPMWIOPKT 22
|||||
RESULT 9
AAM32220
ID      AAM32220 standard; protein; 222 AA.
XX
XX
AC      AAM32220;
XX
XX      03-FEB-1998 (first entry)
XX
XX      Bovine alpha-S2 casein precursor.
XX
XX      Bovine alpha-S2 casein precursor; growth promoting; mitogenic assay;
XX      platelet-derived growth factor; insulin-like derived growth factor;
XX      Bos taurus.
XX
XX      WO9716460-A1.
XX
XX      09-MAY-1997.
XX
XX      31-OCT-1996; 96WO-GB02658.
XX
XX      31-OCT-1995; 95GB-0022302.
XX
XX      (UYLI-) UNITV LIVERPOOL.
XX
XX      Liu Q, Smith JA, Wilkinson MC;
XX
XX      WPI; 1997-272048/24.
XX
XX      Manufacture of medicament or foodstuff for promoting growth - using
XX      peptide(s) with a sequence identical to the C-terminal end of an
XX      alpha-S2 casein precursor
XX
XX      Disclosure; Page 3; 33pp; English.
XX
XX      The present sequence represents bovine alpha-S2 casein precursor.
XX      Peptides having an amino acid sequence which is substantially identical
XX      to the C-terminal end of an alpha-S2 casein precursor, are used for
XX      manufacture of a medicament or foodstuff for promoting growth in humans
XX      or animals.
XX
XX      Sequence 222 AA;
XX
XX      Query Match 100.0%; Score 92; DB 18; Length 222;
XX      Best Local Similarity 100.0%; Pred. No. 2.1e-06;
XX      Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      QY      1 VYOHOKAMKPMWIOPKT 16
XX      |||||
XX      DB      198 VYOHOKAMKPMWIOPKT 213
XX
XX      RESULT 10
XX      AAE17468
XX      ID      AAE17468 standard; Protein; 222 AA.
XX
XX      AAE17468;
XX
XX      22-APR-2002 (first entry)
XX
XX      Bovine alpha-S2 casein precursor protein.
XX
XX      Bovine alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
XX      collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
XX      chewing gum; cosmetic; wrinkling; periodontal disease.
XX
XX      Bos sp.
XX
XX      WO200202133-A2.

```

```

XX      10-JAN-2002.
XX
XX      13-JUN-2001; 2001WO-GB02601.
XX
XX      30-JUN-2000; 2000GB-0016189.
XX
XX      (PEPS-) PEPSYN LTD.
XX
XX      Smith JA;
XX
XX      WPI; 2002-154690/20.
XX
XX      Use of peptide or its derivative containing an amino acid sequence in
XX      alpha-S2 casein precursor in the manufacture of a medicament for
XX      alleviating or preventing periodontal disease and an effect of aging in
XX      skin -
XX
XX      Claim 8, Page 6; 27pp; English.
XX
XX      The invention relates to a composition comprising a peptide or its
XX      derivative. The peptide contains an amino acid sequence from alpha-S2
XX      casein precursor. The peptides stimulate the growth of fibroblasts, and
XX      thus the synthesis and secretion of collagen. The peptides also stimulate
XX      the growth of keratinocytes, which aid in the formation and regeneration
XX      of skin surface. The peptide is useful in the manufacture of a medicament
XX      preventing periodontal disease and a chewing gum, for alleviating or
XX      composition for alleviating or preventing an effect of aging,
XX      particularly wrinkling of the skin. The present sequence is bovine
XX      alpha-S2 casein precursor protein.
XX
XX      Sequence 222 AA;
XX
XX      Query Match 100.0%; Score 92; DB 23; Length 222;
XX      Best Local Similarity 100.0%; Pred. No. 2.1e-06;
XX      Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      QY      1 VYOHOKAMKPMWIOPKT 16
XX      |||||
XX      DB      198 VYOHOKAMKPMWIOPKT 213
XX
XX      RESULT 11
XX      AAM93885
XX      ID      AAM93885 standard; peptide; 26 AA.
XX
XX      AAM93885;
XX
XX      25-JUN-1999 (first entry)
XX
XX      Bifidobacterium bifidus stimulating peptide 21.
XX
XX      Bifidobacterium bifidus
XX
XX      Bifidogenic peptide; protease; treatment; microbe-related disease;
XX      bacteria; fungi; Yeast; Protozoa; virus; mycoplasma; filaria; plasmodia;
XX      infection; inflammation; microbial induced tumor; degenerative disorder;
XX      diarrhoea; colic; oral microflora; intestinal microflora; caries;
XX      vaginal microflora.
XX
XX      Bifidobacterium bifidus.
XX
XX      WO9914231-A2.
XX
XX      25-MAR-1999.
XX
XX      16-SEP-1998; 98WO-EP05899.
XX
XX      11-FEB-1998; 98DE-1005385.
XX
XX      16-SEP-1997; 97DE-1040604.
XX
XX      (FORS/) FORSSMANN W.
XX
XX      Forsmann W, Liepke C, Zucht H;

```

XX WPI; 1995-244022/20.
 XX Milk-derived peptides that stimulate *Bifidobacterium bifidus*
 XX Claim 2; Page 3; 25pp; German.
 XX This invention describes milk-derived bifidogenic peptides and their
 CC active derivatives or fragments, and combinations of them produced by
 CC chemical coupling. Such are produced from bovine or human milk by
 CC treatment for 2 hr with proteases, then centrifuging to remove fat and
 CC acidifying to pH 2 to precipitate proteins. The solution phase is then
 CC subjected to reverse-phase high-performance liquid chromatography (HPLC)
 CC and cation-exchange HPLC, the fractions adjusted to salt content below
 CC 25 mM (by dialysis or reverse-phase HPLC) and tested for activity by
 CC growing *Bifidobacterium bifidus* and *Escherichia coli* in presence of the
 CC fractions. Those fractions for which (Bw-B0) - (Bw-E0) is at least 0.15
 CC are selected where Bw = germ count after 16 hr culture of *B. bifidus* in
 CC 50% Elliker broth containing peptide at 0.2 mg/ml, B0 = germ count under
 CC similar conditions in a peptide-free control, Bw = germ count after 16
 CC hr culture of *E. coli* in 3 g/l tryptic broth containing peptide at
 CC 0.2 mg/ml, E0 = germ count under similar conditions in a peptide-free
 CC control. The peptides AAW3865-W93888 are used to treat microbe-related
 CC diseases caused by bacteria, fungi, yeast, protozoa, viruses,
 CC mycoplasma, filaria and plasmodia, e.g. infections, inflammation, colic,
 CC abnormally induced tumours or degenerative disorders, diarrhoea, colic,
 CC abnormalities in oral, intestinal or vaginal microflora, or carries.

XX Sequence 26 AA;

Query Match 88.6%; Score 81.5; DB 20; Length 26;
 Best Local Similarity 94.1%; Pred. No. 8.8e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 VYOHOKAM-KPMIOPKT 16
 |||||
 DB 1 VYOHOKAMKPMIOPKT 17

RESULT 12

AAE17469
 ID AAE17469 standard; Protein; 223 AA.

XX AAE17469;

XX 22-APR-2002 (first entry)

XX Alpha-S2 casein precursor (alpha-S2-CN) protein #1.

XX Alpha-S2 casein; alpha-S2-CN; dermatological; antiinflammatory; cosmetic;
 KM fibroblast; collagen; keratinocyte; skin regeneration; medicament; aging;
 KM toothpaste; chewing gum; cosmetic; wrinkling; periodontal disease.

XX Unidentified.

XX WO200202133-A2.

XX 10-JAN-2002.

XX 13-JUN-2001; 2001WO-GB02601.

XX 30-JUN-2000; 2000GB-0016189.

XX (PEPS-) PEPSYN LTD.

XX Smith JA;

XX WPI; 2002-154690/20.

XX Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin

XX Disclosure; Page 8; 27pp; English.

XX The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is alpha-S2
 CC casein precursor (alpha-S2-CN) protein.

XX Sequence 223 AA;

Query Match 83.7%; Score 77; DB 23; Length 223;
 Best Local Similarity 87.5%; Pred. No. 0.00045;
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPKT 16
 |||||
 DB 199 VDOHOKAMKPMIOPKT 214

RESULT 13

AAE17470
 ID AAE17470 standard; Protein; 223 AA.

XX AAE17470;

XX 22-APR-2002 (first entry)

XX Goat alpha-S2 casein E precursor protein.

XX Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KM chewing gum; cosmetic; wrinkling; periodontal disease.

XX Capra hircus.

XX WO200202133-A2.

XX 10-JAN-2002.

XX 13-JUN-2001; 2001WO-GB02601.

XX 30-JUN-2000; 2000GB-0016189.

XX (PEPS-) PEPSYN LTD.

XX Smith JA;

XX WPI; 2002-154690/20.

XX Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin

XX Claim 8; Page 8; 27pp; English.

XX The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat

CC alpha-S2 casein E precursor protein
XX
SQ Sequence 223 AA;

Sequence 223 AA;

Query Match	83.7%	Score 77;	DB 23;	length 223;
Best Local Similarity	87.5%;	Pred. No. 0.00045;		
Matches 14; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY	1	VYQHOKAMKPWIQPKT	16
Db	199	VDHQKAMKPWTQPKT	214

RESULT 14
AAE17471
ID AAE17471 standard; Protein; 223 AA
vv

AC AAE17471;

DT 22-APR-2002 (first entry)

DE *Capra hircus* alpha-S2 casein C precursor protein

KM Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging
 KW chewing gum; cosmetic; wrinkling; periodontal disease.

OS Capra hircus.

PN WO200202133-A2

PD 10-JAN-2002.

PF 13-JUN-2001; 2001WO-GB02601.

PR 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD

Smith JA, PI

DR WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in
PT alpha-S2 casein precursor in the manufacture of a medicament for
PT alleviating or preventing periodontal disease and an effect of aging in
PT skin -

PS Claim 8; Page 8; 27pp; English

CC The invention relates to a composition comprising a peptide or its
CC derivative. The peptide contains an amino acid sequence from alpha-S2
CC casein precursor. The peptides stimulate the growth of fibroblasts, and
CC thus the synthesis and secretion of collagen. The peptides also stimulate
CC the growth of keratinocytes, which aid in the formation and regeneration
CC of skin surface. The peptide is useful in the manufacture of a medicament
CC in the form of a toothpaste or a chewing gum, for alleviating or
CC preventing periodontal disease and a medicament in the form of a cosmetic
CC composition for alleviating or preventing an effect of aging,
CC particularly wrinkling of the skin. The present sequence is goat
CC alpha-S2 casein C precursor protein.

SQ Sequence 223 AA

Query Match	83.7%	Score 77;	DB 23;	Length 223;
Best Local Similarity	87.5%;	Pred. No. 0.00045;		
Matches 14; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

```
QY      1 VYQHOKAMKRWIQPKT 16
          | | | | | | | | | |
Db      199 VDQHOKAMKRWIQPKT 214
```

RESULT 15	
AAE17473	
ID	AAE17473 standard; Protein; 223 AA
XY	

AC AA617473
VY

DT 22-APR-2002 (First entry)
 XY

DE Sheep alpha-S2 casein precursor protein
xy

KM Sheep; alpha-82 casein; dermatological; antiinflammatory; fibroblast;
KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging
KM chewing gum; cosmetic; wrinkling; periodontal disease.

OS Ovis sp.
vv

PN WO200202133-A2

PD 10-JAN-2002

PF 13-JUN-2001; 2001WO-GB02601
VV

PR 30-JUN-2000; 2000GB-0016189

PA (PEPS-) PEPSYN LTD.

PI Smith JA;
vv

DR WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in
PT alpha-S2 casein precursor in the manufacture of a medicament for
PT alleviating or preventing periodontal disease and an effect of aging in
PT skin -

PS Claim 8; Page 9; 27pp; English

CC The invention relates to a composition comprising a peptide or its
CC derivative. The peptide contains an amino acid sequence from alpha-S2
CC casein precursor. The peptides stimulate the growth of fibroblasts, and
CC thus the synthesis and secretion of collagen. The peptides also stimulate
CC the growth of keratinocytes, which aid in the formation and regeneration
CC of skin surface. The peptide is useful in the manufacture of a medicament
CC in the form of a toothpaste or a chewing gum, for alleviating or
CC preventing periodontal disease and a medicament in the form of a cosmetic
CC composition for alleviating or preventing an effect of aging,
CC particularly wrinkling of the skin. The present sequence is sheep
CC alpha-S2 casein precursor protein.

Sequence	223 AA;
5Q	

Query Match	83.7%	Score 77;	DB 23;	Length 223;
Best Local Similarity	87.5%	Pred. No. 0.00045;		
Matches 14; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

```

QY      1 VYQHOKAMKRWIQPKT 16
        | | | | | | | | | |
Db      199 VDQHQKAMKRWIQPKT 214

```

Search completed: July 30, 2003, 16:23:26
Job time : 33.5238 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 35.625 Seconds
(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-2

Perfect score: 119
Sequence: 1 VYOHQAKMKPWIOPKTKVPIY 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

.SPREMBL_23:*
1: sp archaea:*
2: sp bacteria:*
3: sp fungi:*
4: sp human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteria:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	86.6	210	6	O62825 bubalus bub
2	95	79.8	223	6	Q9MTU6 Qmyus capra hircu
3	86	72.3	223	6	Q9TUT7 Q9TUT7 capra hircu
4	86	72.3	223	6	Q9GK07 Q9GK07 capra hircu
5	52	43.7	80	10	O81534 saccharum o
6	52	43.7	85	10	O81399 orobancha c
7	52	43.7	340	10	O96568 ipomoea pur
8	52	43.7	379	10	O04065 perilla fru
9	52	43.7	366	10	O23923 digitalis l
10	52	43.7	388	10	O9F837 coretia hyb
11	52	43.7	389	10	Q81P19 nleremergi
12	52	43.7	389	10	O43040 petunia hyb
13	52	43.7	389	10	O93XP8 nicotiana t
14	52	43.7	389	10	O9M5B2 petunia hyb
15	52	43.7	390	10	O9LRB2 scutellaria
16	52	43.7	390	10	O48564 scutellaria

17	52	43.7	390	10	O22122 scutellaria
18	50	42.0	311	5	O19257 scenorrhadi
19	50	42.0	633	11	Q9MW66 Q9MW66 mus musculu
20	50	42.0	1877	3	Q9USR3 Q9USR3 echinosacch
21	49	41.2	282	5	O44074 Q44074 ascaris suu
22	49	41.2	342	16	O8YW06 O8YW06 anabaena sp
23	48	40.3	225	4	O9ETR9 O9ETR9 homo sapien
24	48	40.3	378	16	O8CP60 O8CP60 staphylococ
25	48	40.3	374	10	O81476 O81476 brassica na
26	48	40.3	382	10	O93YX5 O93YX5 vitis sp. c
27	48	40.3	382	10	O9FEK7 O9FEK7 vitis labru
28	48	40.3	382	10	O81P44 O81P44 parthenocis
29	48	40.3	382	10	O9FRW2 O9FRW2 vitis ripar
30	48	40.3	382	10	O9SPW2 O9SPW2 vitis ripar
31	48	40.3	392	10	O81P22 O81P22 parthenocis
32	48	40.3	392	10	O9S982 O9S982 vitis. still
33	48	40.3	382	10	O81P23 O81P23 cistus rhom
34	48	40.3	382	10	O94G58 O94G58 vitis vinif
35	48	40.3	382	10	O944W7 O944W7 vitis sp. c
36	48	40.3	382	10	O944W7 O944W7 vitis sp. c
37	48	40.3	382	10	O93V86 O93V86 humulus lup
38	48	40.3	395	10	O9FR70 O9FR70 allaria pe
39	48	40.3	399	10	O94LW8 O94LW8 humulus lup
40	48	40.3	408	10	O9SLY0 O9SLY0 peliotum nu
41	48	40.3	418	16	O9MYW5 O9MYW5 thermocoga
42	48	40.3	704	4	O9H992 O9H992 homo sapien
43	47	39.5	56	10	O8H250 O8H250 vaccinium m
44	47	39.5	80	10	O9S856 O9S856 glycine max
45	47	39.5	115	10	O9AXX8 O9AXX8 brassica ol

ALIGNMENTS

RESULT 1					
ID	O62825	PRELIMINARY;	PRT;	210	AA.
AC	O62825;				
DT	01-AUG-1998 (TREMblrel. 07, Created)				
DT	01-NOV-1999 (TREMblrel. 12, Last sequence update)				
DT	01-JUN-2001 (TREMblrel. 17, Last annotation update)				
DE	AS2-casein (fragment).				
OS	Bubalus bubalis (Domestic water buffalo).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bubalus.				
OX	NCBI_TaxID=89462;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Mammary gland;				
KA	Dae P., Jain S., Gary L.C.;				
RT	"Cloning and nucleotide sequence of cDNA encoding as2-casein in B.				
RT	bubalis";				
RL	Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.				
DR	EMBL; AJ005431; CAA06534.2; -				
DR	InterPro; IPR001588; Casein.				
DR	Pfam; PF00363; casein; 2.				
FT	NON-TER				
FT	SEQUENCE 210 AA; 24700 MW; 05DEF95963F1132C CRC64;				
QY	Query Match	86.6%;	Score 103;	DB 6;	Length 210;
QY	Best Local Similarity	85.7%;	Pred. No. 5.1e-08;		
QY	Matches 18;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0;
Db	1	VYOHQAKMKPWIOPKTKVPIY 21			
Db	186	VYOHQAKMKPWIOPKTKVPIY 206			
RESULT 2					
ID	O9MYU6	PRELIMINARY;	PRT;	223	AA.
AC	O9MYU6;				

DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE Alpha 82-casein.
 GN CSN152.
 OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OX NCBI_TaxID=9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lagonigro R., Pilla F., Matasino D., Zullo A.;
 RT "A new allele of goat alpha 82-casein gene."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ289716; CAB94236.1; -
 DR InterPro: IPR001588; Casein.
 DR Pfam: PF00363; caseins; 2.
 DR PROSITE: PS00306; CASEIN_ALPHA_BETA; 1.
 SQ SEQUENCE 223 AA; 26403 MW; 0E1F83F24DA8CE2 CRC64;

Query Match 79.8%; Score 95; DB 6; Length 223;
 Best Local Similarity 81.0%; Pred. No. 9.3e-07;
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHKAMKPMIOPTKVIPIY 21
 DB 199 VDOHOKAMKPMIOPTKVIPIY 219

RESULT 3
 Q9TTQ7 PRELIMINARY; PRT; 223 AA.
 AC Q9TTQ7;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE Alpha 82-casein.
 GN CSN152.
 OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OX NCBI_TaxID=9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Velti C.C., Pilla F.F., Lagonigro R.R.;
 RT "A new allele of goat alpha 82-casein."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ249995; CAB59920.1; -
 DR InterPro: IPR001588; Casein.
 DR Pfam: PF00363; caseins; 2.
 DR PROSITE: PS00306; CASEIN_ALPHA_BETA; 1.
 SQ SEQUENCE 223 AA; 26433 MW; CE9FDC8D7688293 CRC64;

Query Match 72.3%; Score 86; DB 6; Length 223;
 Best Local Similarity 76.2%; Pred. No. 2.3e-05;
 Matches 16; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 VYOHKAMKPMIOPTKVIPIY 21
 DB 199 VDOHOKAMKPMIOPTKVIPIY 219

RESULT 4
 Q9GK07 PRELIMINARY; PRT; 223 AA.
 AC Q9GK07;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Alpha 82-casein.
 GN CSN152.

QY 5 QKAMKPMIOPTKVIPIY 18
 DB 33 QKAMKPMIOPTKVIPIY 46

OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OX NCBI_TaxID=9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Velti C.C., Pilla F.F., Lagonigro R.R.;
 RT "A new allele of alpha 82-casein."
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ287310; CAC21704.2; JOINED.
 DR EMBL; AJ287311; CAC21704.2; JOINED.
 DR EMBL; AJ242728; CAC21704.2; JOINED.
 DR EMBL; AJ297312; CAC21704.2; JOINED.
 DR EMBL; AJ297313; CAC21704.2; JOINED.
 DR EMBL; AJ297314; CAC21704.2; JOINED.
 DR EMBL; AJ242527; CAC21704.2; JOINED.
 DR EMBL; AJ287315; CAC21704.2; JOINED.
 DR EMBL; AJ287316; CAC21704.2; JOINED.
 DR EMBL; AJ242526; CAC21704.2; JOINED.
 DR EMBL; AJ242528; CAC21704.2; JOINED.
 DR EMBL; AJ242533; CAC21704.2; JOINED.
 DR InterPro: IPR001588; Casein.
 DR Pfam: PF00363; caseins; 2.
 DR PROSITE: PS00306; CASEIN_ALPHA_BETA; 1.
 SQ SEQUENCE 223 AA; 26432 MW; CE97658D7688C9D CRC64;

Query Match 72.3%; Score 86; DB 6; Length 223;
 Best Local Similarity 76.2%; Pred. No. 2.3e-05;
 Matches 16; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 VYOHKAMKPMIOPTKVIPIY 21
 DB 199 VDOHOKAMKPMIOPTKVIPIY 219

RESULT 5
 O81534 PRELIMINARY; PRT; 80 AA.
 AC O81534;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Chalcone synthase (fragment).
 GN CHS.
 OS Saccharum officinarum (Sugarcan).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACAD clade; Panicoideae; Andropogonaceae; Saccharum.
 OX NCBI_TaxID=4547;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veronesi C., Thalonarn P.;
 RT "Sugarcan resistance to Striga hermonthica."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF079174; AAC2797.1; -
 DR HSSP; P30074; 106F.
 DR InterPro: IPR001099; N-C_synthase.
 DR Pfam: PF00195; Chalcone_synthase; 1.
 DR Pfam: PF000453; N-C_synthase; 1.
 FT NON_TER
 FT NON_TER
 SQ SEQUENCE 80 AA; 8744 MW; 7B94776D2A46F418 CRC64;

Query Match 43.7%; Score 52; DB 10; Length 80;
 Best Local Similarity 64.3%; Pred. No. 1.4;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QKAMKPMIOPTKVIPIY 18
 DB 33 QKAMKPMIOPTKVIPIY 46

RESULT 6

081399

ID 081399 PRELIMINARY; PRT; 85 AA.

AC 081399;

DT 01-NOV-1998 (TREMblrel. 08, Created)

DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)

DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)

DE Chalcone synthase (Fragment).

GN CHS.

OS Orobanchaceae

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; lamids; Lamiales; Orobanchaceae; Orobanchaeae; Orobanchae.

OC NCBI_TaxID=78542;

RN [1]

RP SEQUENCE FROM N.A.

RA Veronesi C., Labrousse P., Thallouarn P.;

RT "Histological, physiological and molecular aspects of sunflower

RL resistance to Orobanchaceae.",

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF074401; AAC26132.1; -.

DR HSSP; P30074; 1D6F.

DR InterPro; IPR001099; N-C synthase.

DR Pfam; PF00195; Chal_still_synC; 1.

DR Prodom; PD000453; N-C synthase; 1.

DR NON TER

FT

SQ SEQUENCE 85 AA; 9483 MW; B9A6ABEDD6D4777E CRC64;

Query Match 43.7%; Score 52; DB 10; Length 85;

Best Local Similarity 64.3%; Pred. No. 1.5;

Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTKV 18

Db 33 OKAIKMGQPKSKI 46

RESULT 7

096568

ID 096568 PRELIMINARY; PRT; 340 AA.

AC 096568;

DT 01-FEB-1997 (TREMblrel. 02, Created)

DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)

DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)

DE Chalcone synthase.

GN CHS-Fil.

OS Ipomoea purpurea (Common morning glory).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; lamids; Solanales; Convolvulaceae; Ipomoea.

OC NCBI_TaxID=4121;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Flower buds;

RA Rauscher M.D., Tiffin P.L., Miller R.E.;

RT "Regulation of anthocyanin gene expression in Ipomoea purpurea.";

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; U74082; ABA41103.1; -.

DR HSSP; P30074; 1D6F.

DR InterPro; IPR001099; N-C synthase.

DR Pfam; PF00195; Chal_still_synC; 1.

DR Pfam; PF02797; Chal_still_synC; 1.

DR Prodom; PD000453; N-C synthase; 1.

DR PROSITE; PS00441; CHALCONE_SYNTH; 1.

SQ SEQUENCE 340 AA; 37519 MW; 06682E6873DCFA39 CRC64;

Query Match 43.7%; Score 52; DB 10; Length 340;

Best Local Similarity 64.3%; Pred. No. 6.2;

Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTKV 18

Db 33 OKAIKMGQPKSKI 46

Db 111 OKAIKMGQPKSKI 124

ID 004065 PRELIMINARY; PRT; 379 AA.

AC 004065;

DT 01-JUL-1997 (TREMblrel. 04, Created)

DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)

DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)

DE Chalcone synthase.

GN CHS.

OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; lamids; Lamiales; Malvaceae; Nepentaceae; Elaeagnaceae;

OC Perilla.

OC NCBI_TaxID=48386;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=96088016; PubMed=9426610;

RX Gong Z., Yamazaki M., Sugiyama M., Tanaka Y., Saito K.;

RA "Cloning and molecular analysis of structural genes involved in

RT anthocyanin biosynthesis and expressed in a forma-specific manner in

RT Perilla frutescens.";

RT Plant Mol. Biol. 35:915-927(1997).

DR EMBL; AB002582; BAA19548.1; -.

DR HSSP; P30074; 1D6F.

DR InterPro; IPR001099; N-C synthase.

DR Pfam; PF00195; Chal_still_synC; 1.

DR Pfam; PF02797; Chal_still_synC; 1.

DR Prodom; PD000453; N-C synthase; 1.

DR PROSITE; PS00441; CHALCONE_SYNTH; 1.

SQ SEQUENCE 379 AA; 41380 MW; 5E149C5530BD106 CRC64;

Query Match 43.7%; Score 52; DB 10; Length 379;

Best Local Similarity 64.3%; Pred. No. 6.9;

Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTKV 18

Db 111 OKAIKMGQPKSKI 124

RESULT 9

023923

ID 023923 PRELIMINARY; PRT; 386 AA.

AC 023923;

DT 01-JAN-1998 (TREMblrel. 05, Created)

DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)

DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)

DE Chalcone synthase (Fragment).

OS Digitalis lanata (Foxglove).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; lamids; Lamiales; Antirrhinaceae; Digitalis.

OC NCBI_TaxID=49450;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Leaf;

RA Theoringer C., Lindemann P., Luckner M.;

RT "Expression of phenylalanine ammonia lyase, chalcone synthase and

RT phytochrome in somatic embryogenesis of Digitalis lanata.";

RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF002526; CAA05512.1; -.

DR HSSP; P30074; 1D6F.

DR InterPro; IPR001099; N-C synthase.

DR Pfam; PF00195; Chal_still_synC; 1.

DR Pfam; PF02797; Chal_still_synC; 1.

DR Prodom; PD000453; N-C synthase; 1.

DR PROSITE; PS00441; CHALCONE_SYNTH; 1.

SQ SEQUENCE 386 AA; 42190 MW; 0A88BF6EDDA82F8F CRC64;

Query Match 43.7%; Score 52; DB 10; Length 386;
 Best Local Similarity 64.3%; Pred. No. 7;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 OKAMKPMIOPRTKV 18
 |||||
 Db 106 OKAIKEWGQPKSKI 119

RESULT 10

Q9FS37 PRELIMINARY; PRT; 389 AA.
 AC Q9FS37;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-OCT-2002 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Chalcone synthase.
 OS Torenia hybrida.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanaceae; Solanaceae; Petunia.
 OC NCBI_TaxID=75807;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Summerwave Blue; TISSUE=petal;
 RA Suzuki K., Xue H., Tanaka Y., Fukuchi M., Mutant M.,
 Murakami Y., Katsumoto Y., Tenda S., Kusumi T.,
 "flower color modifications of Torenia hybrida by cosuppression of
 RT anthocyanin biosynthesis genes."
 RL Mol. Breed. 6:239-246(2000).
 DR EMBL; AB012923; BAB20074.1;
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_still_synth; 1.
 DR Pfam; PF02797; Chal_still_synth; 1.
 DR Prodom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE SYNTH; 1.
 SQ SEQUENCE 389 AA; 42402 MW; BD3A78B6C4A6B0F CRC64;

Query Match 43.7%; Score 52; DB 10; Length 388;
 Best Local Similarity 64.3%; Pred. No. 7.1;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 OKAMKPMIOPRTKV 18
 |||||
 Db 111 OKAIKEWGQPKSKI 124

RESULT 11

Q8LP19 PRELIMINARY; PRT; 389 AA.
 AC Q8LP19;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Chalcone synthase.
 OS Nierembergia sp. NB17.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanaceae; Solanaceae; Nierembergia.
 OC NCBI_TaxID=184877;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NB17;
 RA Ueyama Y., Katsumoto Y., Fukui Y., Ohkawa H., Kusumi T., Tanaka Y.,
 RT "flower color modification of Nierembergia sp. by engineering
 RT flavonoid biosynthetic pathway."
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB078515; BAC10998.1;
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_still_synth; 1.

DR Pfam; PF02797; Chal_still_synth; 1.
 DR Prodom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE SYNTH; 1.
 SQ SEQUENCE 389 AA; 42616 MW; 62DA0C7F407792E7 CRC64;

Query Match 43.7%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 7.1;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 OKAMKPMIOPRTKV 18
 |||||
 Db 111 OKAIKEWGQPKSKI 124

RESULT 12

Q43040 PRELIMINARY; PRT; 389 AA.
 AC Q43040;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Chalcone synthase.
 OS ChS-A.
 OC Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanaceae; Solanaceae; Petunia.
 OC NCBI_TaxID=4102;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=96156377; PubMed=8562849;
 RX Shao L., Li Y., Pan A., Cheng Z., Chen M.;
 RT "Molecular cloning, sequencing, and expression in Escherichia coli of
 RT the chalcone synthase gene."
 RL Chin. J. Biotechnol. 11:131-135(1995).
 DR EMBL; S80857; AAB36038.1;
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_still_synth; 1.
 DR Pfam; PF02797; Chal_still_synth; 1.
 DR Prodom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE SYNTH; 1.
 SQ SEQUENCE 389 AA; 42612 MW; A931CF5B255A0A20 CRC64;

Query Match 43.7%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 7.1;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 OKAMKPMIOPRTKV 18
 |||||
 Db 111 OKAIKEWGQPKSKI 124

RESULT 13

Q93XP8 PRELIMINARY; PRT; 389 AA.
 AC Q93XP8;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Chalcone synthase.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanaceae; Solanaceae; Nicotiana.
 OC NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=flower;
 RA Hu H., Sung H., Su J.;
 RT "Cloning and Expression of a Tobacco Chalcone Synthase Gene."
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF311783; AAK49457.1;

DR InterPro: IPR001099; N-C synthase.
 DR Pfam: PF00195; Chal_syll_synth; 1.
 DR Pfam: PF02797; Chal_syll_synth; 1.
 DR ProDom: PD000453; N-C_synthase; 1.
 DR PROSITE: PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 389 AA; 42563 MW; 96C3D42676E8D44 CRC64;

Query Match 43.7%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 7.1;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QXAKMPWIOPTKV 18
 DB 111 QKAIKMGQPKSKI 124

RESULT 14

Q9MSB2 PRELIMINARY; PRT; 389 AA.

AC Q9MSB2 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE Chalcone synthase (EC 2.3.1.74).
 GN CHS.
 OS Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Solanaceae; Petunia.
 OX NCBI_TaxID=4102;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=violate; TISSUE=Corolla;
 RA Hsu Y.H., Durdan S.F., To K.Y.;
 RT "Cloning of 4 pigment-related cDNAs encoding chalcone synthase,
 RT chalcone isomerase, dihydroflavonol-4-reductase and cytochrome b5 from
 RT Petunia hybrida."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF233638; AAF60297.1; -.
 DR HSSP: P30074; 1D6F.
 DR InterPro: IPR001099; N-C synthase.
 DR Pfam: PF00195; Chal_syll_synth; 1.
 DR Pfam: PF02797; Chal_syll_synth; 1.
 DR ProDom: PD000453; N-C_synthase; 1.
 DR PROSITE: PS00441; CHALCONE_SYNTH; 1.
 KW Acyltransferase; Transferase.
 SQ SEQUENCE 389 AA; 42582 MW; 66135700259D7F0C CRC64;

Query Match 43.7%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 7.1;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QXAKMPWIOPTKV 18
 DB 111 QKAIKMGQPKSKI 124

RESULT 15

Q9LRB2 PRELIMINARY; PRT; 390 AA.

AC Q9LRB2 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE Chalcone synthase.
 GN CHS-C.
 OS Scutellaria baicalensis.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Lamiales; Lamiaceae; Scutellarioidae;
 OC Scutellaria.
 OX NCBI_TaxID=55409;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Morita H., Noguchi H., Akiyama T., Shibuya M., Ebizuka Y.;
 RT "Scutellaria baicalensis chalcone synthase (CHS-C) mRNA, complete
 RT cds."
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB046666; BAB03471.1; -.
 DR HSSP: P30074; 1D6F.
 DR InterPro: IPR001099; N-C synthase.
 DR Pfam: PF00195; Chal_syll_synth; 1.
 DR Pfam: PF02797; Chal_syll_synth; 1.
 DR ProDom: PD000453; N-C_synthase; 1.
 DR PROSITE: PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 390 AA; 42506 MW; 0DA2CB3175DEFF6C CRC64;

Query Match 43.7%; Score 52; DB 10; Length 390;
 Best Local Similarity 64.3%; Pred. No. 7.1;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QXAKMPWIOPTKV 18
 DB 111 QKAIKMGQPKSKI 124

Search completed: July 30, 2003, 16:29:22
 Job time : 37.625 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:10:10 ; Search time 6.875 Seconds
(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-2
Perfect score: 119
Sequence: 1 VYOHOKAMKRWIOPKTKVPIY 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	222	1	CAS2_BOVIN
2	95	79.8	223	1	CAS2_CAPRI
3	95	79.8	223	1	CAS2_SHEEP
4	59.5	50.0	225	1	CAS2_PIG
5	59	49.6	182	1	CAS2_RABIT
6	52	43.7	193	1	CAS2_CAMDR
7	52	43.7	389	1	CHS1_LYCES
8	52	43.7	389	1	CHS2_LYCES
9	52	43.7	389	1	CHS2_SOLTU
10	52	43.7	389	1	CHS2_PETHY
11	52	43.7	389	1	CHS2_SOLTU
12	52	43.7	389	1	CHS2_SOLTU
13	52	43.7	389	1	CHS2_SOLTU
14	52	43.7	389	1	CHS2_PETHY
15	52	43.7	389	1	CHS2_PETHY
16	52	43.7	390	1	CHS2_PETHY
17	52	43.7	390	1	CHS2_PETHY
18	52	43.7	400	1	CHS2_PETHY
19	52	43.7	401	1	CHS2_PETHY
20	52	43.7	401	1	CHS2_PETHY
21	52	43.7	401	1	CHS2_PETHY
22	52	43.7	401	1	CHS2_PETHY
23	52	43.7	401	1	CHS2_PETHY
24	52	43.7	401	1	CHS2_PETHY
25	52	43.7	401	1	CHS2_PETHY
26	52	43.7	401	1	CHS2_PETHY
27	52	43.7	401	1	CHS2_PETHY
28	52	43.7	401	1	CHS2_PETHY
29	52	43.7	401	1	CHS2_PETHY
30	52	43.7	401	1	CHS2_PETHY
31	52	43.7	401	1	CHS2_PETHY
32	52	43.7	401	1	CHS2_PETHY
33	52	43.7	401	1	CHS2_PETHY

34	47	39.5	285	1	CHS6_MEDSA
35	47	39.5	311	1	CHS4_TRISTU
36	47	39.5	368	1	CHS1_SOYBN
37	47	39.5	368	1	CHS2_SOYBN
38	47	39.5	368	1	CHS3_SOYBN
39	47	39.5	368	1	CHS4_SOYBN
40	47	39.5	368	1	CHS5_SOYBN
41	47	39.5	368	1	CHS6_SOYBN
42	47	39.5	368	1	CHS7_SOYBN
43	47	39.5	368	1	CHS8_SOYBN
44	47	39.5	368	1	CHS9_SOYBN
45	47	39.5	368	1	CHS10_SOYBN

ALIGNMENTS

RESULT 1
CAS2_BOVIN STANDARD; PRT; 222 AA.
AC P02633; Q9TR51;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alpha-S2 casein precursor [Contains: Casocidin-I].
OS CSN1S2.
OC Bos taurus (bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88186989; PubMed=2833669;
RA Stewart A.F., Bomsing J., Beattie C.W., Shan F., Willis I.M.,
RA Mackinnon A.G.;
RT "Complete nucleotide sequences of bovine alpha S2- and beta-casein
RT cDNAs: comparisons with related sequences in other species.";
RL Mol. Biol. Evol. 4:231-241(1987).
RN [2]
RP SEQUENCE OF 16-222 (A ALLELE).
RX TISSUE=Milk;
RA MEDLINE=77185633; PubMed=862906;
RA Brignon G., Ribadeau-Dumas B., Mercier J.-C., Pellissier J.-P.,
RA Das B.C.;
RT "Complete amino acid sequence of bovine alphaS2-casein.";
RL FEBS Lett. 76:274-279(1977).
RN [3]
RP PARTIAL SEQUENCE (D ALLELE).
RX TISSUE=Milk;
RA MEDLINE=79239837; PubMed=469044;
RA Grocclaude F., Joudrier P., Mahe M.-F.;
RT "A genetic and biochemical analysis of a polymorphism of bovine alpha
RT S2-casein.";
RL J. Dairy Res. 46:211-213(1979).
RN [4]
RP SEQUENCE OF 165-203, AND CHARACTERIZATION OF CASOCIDIN.
RX TISSUE=Milk;
RA MEDLINE=9600204; PubMed=7556666;
RA Zucht H.-D., Rada M., Adernann K., Weagert H.-J., Forssmann W.-G.;
RT "Caesocidin-I: a casein-alpha s2 derived peptide exhibits antibacterial
RT activity.";
RL FEBS Lett. 372:185-188(1995).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- FUNCTION: CASOCIDIN-I INHIBITS THE GROWTH OF ESCHERICHIA COLI AND
CC STREPTOCOCCUS CARNOBIS.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- MASS SPECTROMETRY: MW=4870; METHOD=Electrospray; RANGE=165-203.
CC -1- POLYMORPHISM: AT LEAST TWO ALLELES EXIST. THE SEQUENCE OF THE A
CC ALLELE IS SHOWN HERE. THE D ALLELE SEQUENCE DIFFERS FROM THAT
CC SHOWN IN HAVING A DELETION OF NINE RESIDUES, WHICH MAY BE 49-58,

CC 50-59, OR 51-60
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC -1- DATABASE: NAME=Protein Spotlight;
 CC NOTE=Issue 16 of November 2001;
 CC WWW="http://www.expasy.org/spotlight/articles/sp101016.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, M16644; AAA30479.1; -
 CC PIR, J02008; KABS02.
 CC InterPro: IPR001588; Casein.
 CC Pfam: PF00363; Casein; 2.
 CC PROSITE, PS00306; CASEIN_ALPHA_BETA, 1.
 CC Milk; Phosphorylation; Signal; Repeat; Antibiotic.
 CC SIGNAL 1 15
 CC CHAIN 16 222 ALPHA-S2 CASEIN.
 CC PEPTIDE 165 203 CASOCIDIN-1.
 CC REPEAT 158 222
 CC MOD_RES 23 23 PHOSPHORYLATION.
 CC MOD_RES 24 24 PHOSPHORYLATION.
 CC MOD_RES 25 25 PHOSPHORYLATION.
 CC MOD_RES 31 31 PHOSPHORYLATION.
 CC MOD_RES 71 71 PHOSPHORYLATION.
 CC MOD_RES 72 72 PHOSPHORYLATION.
 CC MOD_RES 73 73 PHOSPHORYLATION.
 CC MOD_RES 76 76 PHOSPHORYLATION.
 CC MOD_RES 144 144 PHOSPHORYLATION.
 CC MOD_RES 146 146 PHOSPHORYLATION.
 CC CONFLICT 102 102 Q -> E (IN REF. 2).
 CC SEQUENCE 222 AA; 26019 MW; 81E7408AF1C12F7C CRC64;
 CC
 CC Query Match 100.0%; Score 119; DB 1; Length 222;
 CC Best Local Similarity 100.0%; Pred. No. 9, 2e-11;
 CC Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC 1 VYOHOKAMKPMIOPKTVIPY 21
 CC 198 VYOHOKAMKPMIOPKTVIPY 218
 CC
 CC RESULT 2
 CC CAS2_CAPHI STANDARD; PRT; 223 AA.
 CC AC P33049;
 CC DT 01-OCT-1993 (Rel. 27, Created)
 CC DT 01-OCT-1993 (Rel. 27, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Alpha-S2 casein precursor (Alpha-S2-CN).
 CC GN CSNIS2.
 CC OS Capra hircus (Goat).
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC OC Bovidae; Caprinae; Capra.
 CC OX NCBI_TaxID=9925;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=93216130; PubMed=8462880;
 CC RA Bouniol C.; Brignon G.; Mahe M.F.; Prinz C.;
 CC RT "Characterization of goat allelic alpha-s2-caseins A and B: further
 CC RT evidence of the phosphorylation code of caseins.";
 CC RL Protein Seq. Data Anal. 5:213-218(1993).

RN [3]
 RP SEQUENCE FROM N.A. (VARIANT C).
 RX MEDLINE=95030556; PubMed=7943951;
 RA Bouniol C.; Brignon G.; Mahe M.F.; Prinz C.;
 RT "Biochemical and genetic analysis of variant C of caprine alpha s2-
 RT casein (Capra hircus).";
 RL Ann. Genet. 25:173-177(1994).
 CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
 CC CALCIUM PHOSPHATE
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
 CC -1- POLYMORPHISM: THREE ALLELES OF ALPHA-S2 CASEIN ARE KNOWN: A, B AND
 CC C. THE FREQUENCIES OF THE ALLELES IS ESTIMATED TO BE 0.85, 0.04
 CC AND 0.11 IN THE FRENCH DAIRY BREEDS 'ALPINE' AND 'SAENEN'.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, X65160; CAA46278.1; -
 CC DR EMBL, S74171; AAB32166.1; -
 CC PIR, I46995; I46995.
 CC DR PIR, S33881; UN0547.
 CC InterPro: IPR001588; Casein.
 CC Pfam: PF00363; Caseins; 2.
 CC PROSITE, PS00306; CASEIN_ALPHA_BETA, 1.
 CC Milk; Phosphorylation; Signal; Repeat; Polymorphism.
 CC SIGNAL 1 15
 CC CHAIN 16 223 ALPHA-S2 CASEIN.
 CC REPEAT 77 141
 CC MOD_RES 159 223
 CC MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 77 77 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 145 145 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 147 147 PHOSPHORYLATION (BY SIMILARITY).
 CC VARIANT 79 79 E -> K (IN VARIANT B).
 CC VARIANT 182 182 K -> I (IN VARIANT C).
 CC SEQUENCE 223 AA; 26389 MW; 187DEFA42FD688291 CRC64;
 CC
 CC Query Match 79.8%; Score 95; DB 1; Length 223;
 CC Best Local Similarity 81.0%; Pred. No. 3, 6e-07;
 CC Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 CC
 CC 1 VYOHOKAMKPMIOPKTVIPY 21
 CC 199 VYOHOKAMKPMIOPKTVIPY 219
 CC
 CC RESULT 3
 CC CAS2_SHEEP STANDARD; PRT; 223 AA.
 CC AC P04654;
 CC DT 13-AUG-1987 (Rel. 05, Created)
 CC DT 13-AUG-1987 (Rel. 05, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Alpha-S2 casein precursor.
 CC GN CSNIS2.
 CC OS Ovis aries (Sheep).
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC OC Bovidae; Caprinae; Ovis.
 CC OX NCBI_TaxID=9940;
 CC RN [1]

```

RP SEQUENCE FROM N.A.
RX MEDLINE=86104467; PubMed=3002499;
RA Boismard M., Petrisant G.;
RT "Complete sequence of ovine alpha s2-casein messenger RNA.";
RL Biochimie 67:1043-1051(1985).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X03238; CA26983.1; -.
DR PIR; A25070; KASHS2.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN ALPHA BETA; 1.
KW Milk; Phosphorylation; Repeat; Signal.
FT SIGNAL 1 15
FT CHAIN 1 223 ALPHA-S2 CASEIN.
FT REPEAT 77 141
FT REPEAT 159 223
FT MOD_RES 23 23 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 24 24 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 25 25 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 72 72 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 73 73 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 74 74 PHOSPHORYLATION (POTENTIAL).
FT VARIANT 64 64 D -> N.
SQ SEQUENCE 223 AA; 26332 MW; 67212935E27426D7 CRC64;

Query Match 79.8%; Score 95; DB 1; Length 223;
Best Local Similarity 81.0%; Pred. No. 3.6e-07;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHOKMKRPIQPKTVITY 21
Db 199 VDOHOKMKRPIQPKTVITY 219

RESULT 4
ID CAS2_PIG STANDARD; PRT; 235 AA.
AC P39936;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-s2 casein precursor.
OS CSNIS2.
OC Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9923;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92367960; PubMed=1503276;
RA Alexander L.J., Das Gupta N.A., Beattie C.W.;
RT "The sequence of porcine alpha s2-casein cDNA.";
RL Anim. Genet. 23:365-367(1992).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
-----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X54975; CA38719.1; -.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN ALPHA BETA; 1.
KW Milk; Phosphorylation; Signal.
FT SIGNAL 1 15
FT CHAIN 1 235 ALPHA-S2 CASEIN.
FT MOD_RES 16 235 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 70 70 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 71 71 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 235 AA; 27570 MW; C903B760D184C14C CRC64;

Query Match 50.0%; Score 59.5; DB 1; Length 235;
Best Local Similarity 56.5%; Pred. No. 0.08;
Matches 13; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

QY 1 VYOHOKMKRPIQPKTVITY 20
Db 208 VYOHOKMKRPIQPKTVITY 230

RESULT 5
ID CAS3_RABIT STANDARD; PRT; 182 AA.
AC P50419;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Alpha-S2B casein precursor.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=New Zealand white; TISSUE=Mammary gland;
RX MEDLINE=94107245; PubMed=8280077;
RA Dawson S.P., Wilde C.J., Tighe P.J., Mayer R.J.;
RT "Characterization of two novel casein transcripts in rabbit mammary
RT gland.";
RL Biochem. J. 296:777-784(1993).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X76909; CA45423.1; -.
DR PIR; S39776; S39776.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
-----

```

DR PROSITE: P800306; CASEIN_ALPHA_BETA; 1.
 KW Milk; Phosphorylation; Signal.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 16 182 ALPHA-S2B CASEIN.
 SQ SEQUENCE 182 AA; 21597 MW; F2433C2DA4550FC CRC64;
 Query Match 49.6%; Score 59; DB 1; Length 182;
 Best Local Similarity 42.9%; Pred. No. 0.074;
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 VYQHQKAMKPMIOPTKYIPY 21
 :||:|||||:|||||:
 DB 83 LYQPTVMDPMWTRAEKTRAIIP 103

RESULT 6
 ID CAS2_CAMDR STANDARD; PRT; 193 AA.
 AC 097934;
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Alpha-S2 casein precursor.
 CS N152.
 OS Camelus dromedarius (Dromedary) (Arabian camel).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
 OX NCBI_TaxId=9838;
 RN (1)
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN=Somali; TISSUE=Udder;
 RX MEDLINE=98291310; PubMed=9627840;
 RA Kappeler S., Farah Z., Puhon Z.;
 RT "Sequence analysis of Camelus dromedarius milk caseins";
 RL J. Dairy Res. 65:209-222(1998).
 CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
 CC -1- CALCIUM PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: AJ012629; CA110078.1; -
 CC DR InterPro: IPR001588; Casein.
 CC DR Pfam: PF00363; caseins.2.
 CC DR PROSITE: P800306; CASEIN_ALPHA_BETA; FALSE_NEG.
 CC KW Milk; Phosphorylation; Signal.
 CC FT SIGNAL 1 15
 CC FT CHAIN 16 193 ALPHA-S2 CASEIN.
 CC SQ SEQUENCE 193 AA; 22564 MW; 2843256F8FD2ED13 CRC64;
 Query Match 43.7%; Score 52; DB 1; Length 193;
 Best Local Similarity 42.9%; Pred. No. 0.87;
 Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 VYQHQKAMKPMIOPTKYIPY 21
 :||:|||||:|||||:
 DB 93 LHQGOIVNPMWDOGKTRAYPR 113

RESULT 7
 ID CHS1_LYCSES STANDARD; PRT; 389 AA.
 AC P23418;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DR Chalcone synthase 1 (EC 2.3.1.74) (Naringenin-chalcone synthase 1).
 GN CHS1.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxId=4081;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cotyledon, Hypocotyl, and Leaf;
 RX MEDLINE=91117196; PubMed=1980524;
 RA O'Neill S.D., Tong Y., Spoerlein B., Forkmann G., Yoder J.I.;
 RT "Molecular genetic analysis of chalcone synthase in Lycopersicon
 RT esculentum and an anthocyanin-deficient mutant.";
 RL Mol. Gen. Genet. 224:279-286(1990).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4'',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: X55194; CA38980.1; -
 CC DR InterPro: IPR001099; N-C_synthase.
 CC DR Pfam: PF00195; Chal_scl1_synth. 1.
 CC DR Pfam: PF02797; Chal_scl1_synthC. 1.
 CC DR ProDom: PD000453; N-C synthase; 1.
 CC DR PROSITE: P800441; CHALCONE SYNTH. 1.
 CC KW Flavonoid biosynthesis; Transferase; Acyltransferase;
 CC KW Multigene family.
 CC FT ACT_SITE 164 164 BY SIMILARITY.
 CC FT SEQUENCE 389 AA; 42552 MW; 553DC695EA96A8B CRC64;
 Query Match 43.7%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 1.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QKAMKPMIOPTKYIPY 18
 :||:|||||:|||||:
 DB 111 QKAIKMGQPSKX 124

RESULT 8
 ID CHS2_LYCSES STANDARD; PRT; 389 AA.
 AC P23419;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2).
 GN CHS2.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxId=4081;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cotyledon, Hypocotyl, and Leaf;
 RX MEDLINE=91117196; PubMed=1980524;

Query Match 43.7%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 1.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 OKANKPMIOPKTV 18
 |||:|||||:
 Db 111 OKAIKEWGQPKSKI 124

RESULT 13

CHSE_IPOPU STANDARD; PRT; 389 AA.
 AC 022047;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chalcone synthase E (EC 2.3.1.74) (Naringenin-chalcone synthase E)
 GN (CHS-E).
 OS Ipomoea purpurea (Common morning glory).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Convolvulaceae; Ipomoea.
 OC NCBI_TaxID=4121;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. FR-35; TISSUE=flower buds;
 RX MEDLINE=97393496; PubMed=9249990;
 RA Fukuda-Tanaka S., Hoshino A., Hisatsomi Y., Habu Y., Hasebe M.,
 RA Iida S.;
 RT "Identification of new chalcone synthase genes for flower pigmentation
 in the Japanese and common morning glories.";
 RL Plant Cell Physiol. 38:754-758(1997)
 CC -1- FUNCTION: The primary product of this enzyme is 4',2',4',6'-
 tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 which can under specific conditions spontaneously isomerize into
 naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 FLAVONOIDS. A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by, and for commercial
 entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 or send an email to license@isb-sdb.ch).
 CC
 CC EMBL; AB001827; BAA21789.1; -.
 DR PIR; T07799; T07799.
 DR InterPro; IPR001099; N-C synthase.
 DR Pfam; PF00195; Chal_still_synthc.1.
 DR Pfam; PF02797; Chal_still_synthc.1.
 DR ProDom; PD000453; N-C synthase.1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 DR Flavonoid biosynthesis; Transferase; Acyltransferase;
 KW Multigene family.
 KW ACT SITE 164
 FT ACT SITE 164 42677 MW; 0CC81302CCBA56A CRC64;
 SQ SEQUENCE 389 AA; 42677 MW; 0CC81302CCBA56A CRC64;

Query Match 43.7%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 1.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 OKANKPMIOPKTV 18
 |||:|||||:
 Db 111 OKAIKEWGQPKSKI 124

Query Match 43.7%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 1.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 OKANKPMIOPKTV 18
 |||:|||||:
 Db 111 OKAIKEWGQPKSKI 124

Qy 5 OKANKPMIOPKTV 18
 |||:|||||:
 Db 111 OKAIKEWGQPKSKI 124

RESULT 14

CHSE_PHANT STANDARD; PRT; 389 AA.
 AC 022046;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chalcone synthase E (EC 2.3.1.74) (Naringenin-chalcone synthase E)
 GN (CHS-E).
 OS Pharbitis nil (Violet) (Japanese morning glory).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Convolvulaceae; Ipomoea.
 OC NCBI_TaxID=35883;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. KK/ZSK-2; TISSUE=flower buds;
 RX MEDLINE=97393496; PubMed=9249990;
 RA Fukuda-Tanaka S., Hoshino A., Hisatsomi Y., Habu Y., Hasebe M.,
 RA Iida S.;
 RT "Identification of new chalcone synthase genes for flower pigmentation
 in the Japanese and common morning glories.";
 RL Plant Cell Physiol. 38:754-758(1997)
 CC -1- FUNCTION: The primary product of this enzyme is 4',2',4',6'-
 tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 which can under specific conditions spontaneously isomerize into
 naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 FLAVONOIDS. A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by, and for commercial
 entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 or send an email to license@isb-sdb.ch).
 CC
 CC EMBL; AB001819; BAA21788.1; -.
 DR InterPro; IPR001099; N-C synthase.
 DR Pfam; PF00195; Chal_still_synthc.1.
 DR Pfam; PF02797; Chal_still_synthc.1.
 DR ProDom; PD000453; N-C synthase.1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 DR Flavonoid biosynthesis; Transferase; Acyltransferase;
 KW Multigene family.
 KW ACT SITE 164
 FT ACT SITE 164 42685 MW; 0FC0E014B9CC0312 CRC64;
 SQ SEQUENCE 389 AA; 42685 MW; 0FC0E014B9CC0312 CRC64;

Query Match 43.7%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 1.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 OKANKPMIOPKTV 18
 |||:|||||:
 Db 111 OKAIKEWGQPKSKI 124

RESULT 15

CHSJ_PETHY STANDARD; PRT; 389 AA.
 AC P22928;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chalcone synthase J (EC 2.3.1.74) (Naringenin-chalcone synthase J).
 GN CHSJ.

```

OS Petunia hybrida (Petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Lamiales; Solanales; Solanaceae; Petunia.
OX NCBI_TaxId=4102;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Violet 30; TISSUE=leaf;
RX MEDLINE=90034197; PubMed=2806915;
RA Koes R.E., Spelt C.E., van den Bizen P.J.M., Mol J.N.M.;
RT "Cloning and molecular characterization of the chalcone synthase
  multigene family of Petunia hybrida.";
  Gene 81:245-257(1989).
  [2]
RN [2]
RP SEQUENCE OF 71-389 FROM N.A.
RC STRAIN=cv. White 137; TISSUE=anther;
RA van Tunen A.J.;
RT Submitted (Mar-1989) to the EMBL/GenBank/DDJ databases.
CC -! FUNCTION: The primary product of this enzyme is 4,2',4',6'-
  tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
  which can under specific conditions spontaneously isomerize into
  naringenin.
CC -! CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
  naringenin-chalcone + 3 CO(2).
CC -! PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
  FLAVONOIDS. A LARGE CLASS OF SECONDARY PLANT METABOLITES. MANY
  OF WHICH ARE BRIGHTLY COLORED.
CC -! SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X14597; CAA32737.1; -.
DR EMBL; X14599; CAA32739.1; -.
DR PIR; D72821; SYPUCJ.
DR PIR; S18136; S18136.
DR InterPro; IPR001099; N-C_synthase.
DR Pfam; PF00195; Chal_stil_synct; 1.
DR PROSITE; PS00797; Chal_stil_synct; 1.
DR PROSITE; PS00453; N-C_synthase; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase;
  Multigene family.
KW ACT SITE
FT ACT SITE 164
FT CONFLICT 75 E -> V (IN REF. 2).
SQ SEQUENCE 389 AA; 42558 MW; F2B3CDBD82E6FDE7D CRC64;

Query Match 43.7%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 1.8;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPTKV 18
DB 111 OKAIKMGOPKSKI 124

```

Search completed: July 30, 2003, 16:24:28
 Job time : 6.875 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: July 30, 2003, 16:16:55 ; Search time 13.375 Seconds
(without alignments)
150.994 Million cell updates/sec

Title: US-09-787-070-2

Perfect score: 119

Sequence: 1 VYOHQAKMKRWIQPKTKVLPY 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR_76:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	222	1 KABOS2	alpha-s2-casein pr
2	95	79.8	223	2 KASHS2	alpha-s2-casein pr
3	95	79.8	223	2 JN0547	alpha-s2-casein pr
4	95	79.8	223	2 I46995	alpha-s2-casein C
5	59.5	50.0	235	2 A48383	alpha-s2-casein -
6	59	49.6	192	2 S39776	alpha-s2-casein fo
7	52	43.7	319	2 S18136	naringenin-chalcon
8	52	43.7	340	2 T10957	naringenin-chalcon
9	52	43.7	389	1 SYRUCJ	naringenin-chalcon
10	52	43.7	389	1 SYRUCN	naringenin-chalcon
11	52	43.7	389	1 SYRUCN	naringenin-chalcon
12	52	43.7	389	2 JCS136	naringenin-chalcon
13	52	43.7	389	2 T07799	naringenin-chalcon
14	52	43.7	390	1 SYSKCD	naringenin-chalcon
15	52	43.7	390	1 SYSKCD	naringenin-chalcon
16	51	42.9	180	2 S39775	naringenin-chalcon
17	50	42.0	311	2 T15977	alpha-s2-casein fo
18	50	42.0	1877	2 T40550	hypothetical prote
19	49	41.2	188	2 UC4680	probable phosphati
20	49	41.2	207	2 UC4679	vascular endotheli
21	49	41.2	342	2 AD2032	hypothetical prote
22	49	41.2	398	2 S16275	naringenin-chalcon
23	49	41.2	400	1 SYZMCC	naringenin-chalcon
24	48	40.3	392	2 S16206	stilbene synthase
25	48	40.3	392	2 S11044	stilbene synthase
26	48	40.3	392	2 S53313	stilbene synthase
27	48	40.3	392	2 S53314	stilbene synthase
28	47	39.5	53	2 F72370	phoH-related prote
29	47	39.5	53	2 S41957	naringenin-chalcon
			128	2 S33611	naringenin-chalcon

30	47	39.5	158	2 B86169	hypothetical prote
31	47	39.5	285	2 S44369	naringenin-chalcon
32	47	39.5	331	2 Q01071	naringenin-chalcon
33	47	39.5	370	2 S44368	naringenin-chalcon
34	47	39.5	383	2 S35165	naringenin-chalcon
35	47	39.5	388	1 SYSYCN	naringenin-chalcon
36	47	39.5	388	1 SYSYCN	naringenin-chalcon
37	47	39.5	388	1 SYSYCN	naringenin-chalcon
38	47	39.5	388	1 SYSYCN	naringenin-chalcon
39	47	39.5	388	2 JQ2249	naringenin-chalcon
40	47	39.5	388	2 JQ2259	naringenin-chalcon
41	47	39.5	388	2 S60472	naringenin-chalcon
42	47	39.5	388	2 JCS516	naringenin-chalcon
43	47	39.5	389	1 SYRUCP	naringenin-chalcon
44	47	39.5	389	1 SYRUCF	naringenin-chalcon
45	47	39.5	389	2 S20932	naringenin-chalcon

ALIGNMENTS

RESULT 1
KABOS2
alpha-s2-casein precursor - bovine
C.Species: Bos primigenius taurus (cattle)
C.Date: 01-Sep-1991 #sequence revision 03-Feb-1994 #text_change 22-Jun-1999
C.Accession: JQ2008; A25087; A91438; S66626; A03107
R.Groenen, M.A.M.; Dijkhof, R.J.M.; Verstege, A.J.M.; van der Poel, J.J.
Gene 123, 187-193, 1993
A.Title: The complete sequence of the gene encoding bovine alpha2-casein.
A.Reference number: JQ2008; MWID:93154583; PMID:8428658
A.Accession: JQ2008
A.Status: translation not shown
A.Molecule type: DNA
A.Residues: 1-222 <GRO>
A.CROSS-references: GB:M94327
R.Stewart, A.F.; Bomsling, J.; Beattie, C.W.; Shah, F.; Willis, I.M.; Mackinlay, A.G.
Mol. Biol. Evol. 4, 231-241, 1987
A.Title: Complete nucleotide sequences of bovine alpha-s2- and beta-casein cDNAs: compar
A.Reference number: A93062; MWID:88188989; PMID:2833669
A.Accession: A29087
A.Status: translation not shown
A.Molecule type: mRNA
A.Residues: 1-222 <STE>
A.CROSS-references: GB:M16644; NID:G162928; PID:AAA0479.1; PID:G162929
R.Brignon, G.; Ribadeau Dumas, B.; Mercier, J.C.; Pelissier, J.P.; Das, B.C.
FEBS Lett. 76, 274-279, 1977
A.Title: Complete amino acid sequence of bovine alpha-s2-casein.
A.Reference number: A91438; MWID:77185633; PMID:862906
A.Contents: A allele
A.Accession: A91438
A.Molecule type: Protein
A.Residues: 16-101, 'EE', 104-222

A.Note: four fractions, previously designated s2, s3, s4, and s6, appear to have the sam
these
J.Grosclaude, F.; Joudrier, P.; Mahe, M.F.
J. Dairy Res. 46, 211-213, 1979
A.Title: A genetic and biochemical analysis of a polymorphism of bovine alpha-s2-casein
A.Reference number: A92771; MWID:79239837; PMID:469044
A.Contents: annotation; D allele
A.Note: the sequence of the D allele has a deletion of nine residues, which may be 49-58
R.Zucht, H.D.; Raida, M.; Adermann, K.; Maegert, H.J.; Forssmann, W.G.
FEBS Lett. 372, 185-188, 1995
A.Title: Casocidin-i, a casein-alpha(s2) derived peptide exhibits antibacterial activity
A.Reference number: S66626; MWID:96000204; PMID:755666
A.Accession: S66626
A.Molecule type: Protein
A.Residues: 165-203 <ZUC>
C.Comment: The sequence of the A allele is shown.
C.Genetics:
A.Gene: alpha2ca
A.Map position: 6
A.Introns: 17/3; 26/3; 33/3; 47/3; 56/3; 65/3; 74/3; 82/3; 97/3; 128/3; 147/3; 156/3; 16

C:Superfamily: alpha-s2-casein
 C:Keywords: mammary gland; milk; phosphoprotein
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-222/Product: alpha-s2-casein #status experimental <MAT>
 F:23,24,25,31,71,72,73,76,144,146,158/Binding site: phosphate (Ser) (covalent) #status P

Query Match 100.0%; Score 119; DB 1; Length 222;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKMKPWTQPKTKVTPY 21
 |||||
 DB 199 VDOHQKMKPWTQPKTKVTPY 218

RESULT 2

KASHS2
 alpha-s2-casein precursor - sheep
 C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C>Date: 31-Mar-1988 #sequence_revansion 31-Mar-1988 #text_change 22-Jun-1999
 C:Accession: A25070; S17856
 R:Boulmiard, M.; Pectrisant, G.
 Biochimie 67, 1043-1051, 1985
 A:Title: Complete sequence of ovine alpha-s2-casein messenger RNA.
 A:Reference number: A25070; MUID:66104467; PMID:3002499
 A:Accession: A25070

A:Molecule type: mRNA
 A:Residues: 1-223 <BO1>
 A:Cross-references: GB:X0328; NID:g1238; PID:CAA2683.1; PID:g732894
 A>Note: 64-Asn was also found
 R:Boulmiard, M.; Hue, D.; Boulmiard, C.; Mercier, J.C.; Gaye, P.
 Eur. J. Biochem. 201, 633-641, 1991
 A:Title: Multiple mRNA species code for two non-allelic forms of ovine alpha-s2-casein.
 A:Reference number: S17856; MUID:9203619; PMID:1335859

A:Accession: S17856
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-93, 'T', 95-223 <BO2>
 C:Superfamily: alpha-s2-casein
 C:Keywords: mammary gland; milk; phosphoprotein
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-223/Product: alpha-s2-casein #status predicted <KA2>
 F:23,24,25,32,55,72,73,74,77,145,147,159/Binding site: phosphate (Ser) (covalent) #status predicted
 F:53,68,146,154,170,198/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 79.8%; Score 95; DB 1; Length 223;
 Best Local Similarity 81.0%; Pred. No. 5.5e-07;
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKMKPWTQPKTKVTPY 21
 |||||
 DB 199 VDOHQKMKPWTQPKTKVTPY 219

RESULT 3

JN0547
 alpha-s2-casein precursor - goat
 C:Species: Capra aegagrus hircus (domestic goat)
 C>Date: 31-Dec-1993 #sequence_revansion 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: S33881; S33880; JN0547; S20620
 R:Boulmiard, C.; Brignon, G.; Mahe, M.F.; Printz, C.
 Protein Seq. Data Anal. 5, 213-218, 1993
 A:Title: Characterization of goat allelic alpha-s2-caseins A and B: further evidence of

A:Reference number: S33880
 A:Accession: S33881
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-223 <BO1>

A>Note: allele B
 A:Accession: S33880
 A:Molecule type: protein
 A:Residues: 16-78, 'E', 80-223 <BO2>
 A:Title: sequence deduced from compositional analysis of peptides

R:Boulmiard, C.
 Gene 125, 235-236, 1993
 A:Title: Sequence of the goat alpha-s2-casein-encoding cDNA.
 A:Reference number: JN0547; MUID:93216130; PMID:8462880
 A:Accession: JN0547

A:Molecule type: mRNA
 A:Residues: 1-78, 'E', 80-223 <BO3>
 A:Cross-references: EMBL:X65160; NID:g955; PID:CAA46278.1; PID:g956

C:Superfamily: alpha-s2-casein
 C:Keywords: mammary gland; milk; phosphoprotein
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-223/Product: alpha-s2-casein #status experimental <MAT>
 F:23,24,25,55,72,73,74,77,145,147,159/Binding site: phosphate (Ser) (covalent) #status predicted
 F:53,146/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 79.8%; Score 95; DB 2; Length 223;
 Best Local Similarity 81.0%; Pred. No. 5.5e-07;
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKMKPWTQPKTKVTPY 21
 |||||
 DB 199 VDOHQKMKPWTQPKTKVTPY 219

RESULT 4

I4695
 alpha s2-casein C - goat
 C:Species: Capra aegagrus hircus (domestic goat)
 C>Date: 21-Feb-1997 #sequence_revansion 21-Feb-1997 #text_change 13-Aug-1999
 C:Accession: I4695
 R:Boulmiard, C.; Brignon, G.; Mahe, M.F.; Printz, C.
 Anim. Genet. 25, 173-177, 1994

A:Title: Biochemical and genetic analysis of variant C of caprine alpha s2-casein (Capr).
 A:Reference number: I4695; MUID:95030556; PMID:7943951
 A:Accession: I4695
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-223 <BO1>
 A:Cross-references: GB:574171; NID:g707033; PID:AA832166.1; PID:g707034
 C:Superfamily: alpha-s2-casein

Query Match 79.8%; Score 95; DB 2; Length 223;
 Best Local Similarity 81.0%; Pred. No. 5.5e-07;
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKMKPWTQPKTKVTPY 21
 |||||
 DB 199 VDOHQKMKPWTQPKTKVTPY 219

RESULT 5

A48383
 alpha s2-casein - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C>Date: 19-Nov-1993 #sequence_revansion 18-Nov-1994 #text_change 03-May-1996
 C:Accession: A48383
 R:Alexander, L.J.; Das Gupta, N.A.; Beattie, C.W.
 Anim. Genet. 23, 365-367, 1992

A:Title: The sequence of porcine alpha s2-casein cDNA.
 A:Reference number: A48383; MUID:92367960; PMID:1503276
 A:Accession: A48383
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-235 <ALE>
 A:Experimental source: mammary gland
 A>Note: sequence inconsistent with the nucleotide translation
 A>Note: sequence extracted from NCBI backbone (NCBIN110884; NCBI:P110885)
 C:Superfamily: alpha-s2-casein

Query Match 50.0%; Score 59.5; DB 2; Length 235;
 Best Local Similarity 56.8%; Pred. No. 0.15;
 Matches 13; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

OY 1 VYOHQKMKPWIOPTKTV---KVTP 20
 DB 208 VHOQKMKMKPWNIKTNSYQIIP 230

RESULT 6

alpha-s2-casein form b precursor - rabbit
 S39776
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 13-Aug-1999
 C:Accession: S39776
 R:Dawson, S.P.; Wilde, C.J.; Tighe, P.J.; Mayer, R.J.
 Biochem. J. 296, 777-784, 1993
 A:Title: Characterization of two novel casein transcripts in rabbit mammary gland.
 A:Reference number: S39775; MID:94107245; PMID:8280077
 A:Accession: S39776
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-182 <DAW>
 A:Cross-references: EMBL:X76909; NID:g439527; PIDN:CAA54231.1; PID:g439528
 C:Superfamily: alpha-s2-casein

Query Match 49.6%; Score 59; DB 2; Length 182;
 Best Local Similarity 42.9%; Pred. No. 0.13;
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 VYOHQKMKPWIOPTKTVIPY 21
 DB 83 LYQYPTVMDPTRAETKAIIP 103

RESULT 7

naringenin-chalcone synthase (EC 2.3.1.74) - garden petunia
 S18136
 C:Species: Petunia x hybrida (garden petunia)
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-May-2000
 C:Accession: S18136
 R:van Tunen, A.J.
 submitted to the EMBL Data Library, March 1989
 A:Reference number: S18136
 A:Accession: S18136
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-319 <TUN>
 A:Cross-references: EMBL:X14599; NID:g20537; PIDN:CAA32739.1; PID:g20538
 C:Superfamily: chalcone synthase
 C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 43.7%; Score 52; DB 2; Length 319;
 Best Local Similarity 64.3%; Pred. No. 2.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 OKAMKPWIOPTKTV 18
 DB 41 OKAIKMGQPKSKI 54

RESULT 8

naringenin-chalcone synthase (EC 2.3.1.74) CHS-FLI - common morning-glory
 T10957
 C:Species: Ipomoea purpurea (common morning-glory)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 05-May-2000
 C:Accession: T10957
 R:Rauscher, M.D.; Tiffin, P.L.; Miller, R.R.
 submitted to the EMBL Data Library, January 1997
 A:Description: Regulation of anthocyanin gene expression in Ipomoea purpurea.
 A:Reference number: Z17229
 A:Accession: T10957
 A:Status: preliminary;
 A:Molecule type: mRNA
 A:Residues: 1-340 <RAU>
 A:Cross-references: EMBL:U74082; NID:g1658119; PID:g1658120

A:Experimental source: flower bud; WMAA genotype (dark flowered)
 C:Genetics:
 A:Gene: CHS-FLI
 C:Superfamily: chalcone synthase
 C:Keywords: acyltransferase; coenzyme A

Query Match 43.7%; Score 52; DB 2; Length 340;
 Best Local Similarity 64.3%; Pred. No. 3;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 OKAMKPWIOPTKTV 18
 DB 111 OKAIKMGQPKSKI 124

RESULT 9

naringenin-chalcone synthase (EC 2.3.1.74) J - garden petunia
 SYRUCI
 N:Alternate names: chalcone synthase
 C:Species: Petunia x hybrida (garden petunia)
 C:Date: 30-Sep-1991 #sequence_revision 17-Mar-2000 #text_change 05-May-2000
 C:Accession: D72821; JS0309
 R:Koes, R.E.; Spelt, C.E.; van den Elzen, P.J.M.; Mol, J.N.M.
 Gene 81, 245-257, 1989
 A:Title: Cloning and molecular characterization of the chalcone synthase multigene family
 A:Reference number: JS0308; MID:90034197; PMID:2806915
 A:Accession: D72821
 A:Molecule type: DNA
 A:Residues: 1-389 <KOE>
 A:Cross-references: EMBL:X14597; NID:g20535; PIDN:CAA32737.1; PID:g20536
 A:Experimental source: strain Violet 30, leaf
 A:Accession: JS0309
 A:Molecule type: DNA
 A:Residues: 1-50, 'D', 52-74, 'V', 76-228, 'I', 230-297, 'L', 299-389 <KO2>
 A:Note: The sequence is revised in Genbank entry EPHSD release 114, (PIDN:CAA32737.1)
 C:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavon
 C:Genetics:
 A:Gene: chad
 A:Map position: V
 A:Note: chad is expressed in various floral tissues and UV illuminated seedlings
 C:Superfamily: chalcone synthase
 C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 43.7%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 3.4;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 OKAMKPWIOPTKTV 18
 DB 111 OKAIKMGQPKSKI 124

RESULT 10

naringenin-chalcone synthase (EC 2.3.1.74) R - garden petunia
 SYRUCN
 N:Alternate names: chalcone synthase
 C:Species: Petunia x hybrida (garden petunia)
 C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000
 C:Accession: A23643
 R:Koes, R.E.; Spelt, C.E.; Reif, H.J.; van den Elzen, P.J.M.; Veltkamp, E.; Mol, J.N.M.
 Nucleic Acids Res. 14, 5229-5239, 1986
 A:Title: Floral tissue of Petunia hybrida (V30) expresses only one member of the chalcon
 A:Reference number: A23643; MID:86286540; PMID:3016642
 A:Accession: A23643
 A:Molecule type: mRNA
 A:Residues: 1-389 <KOE>
 A:Cross-references: GB:X04080; NID:g20541; PIDN:CAA27718.1; PID:g20542
 A:Experimental source: strain Violet 30, flowers
 C:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavon
 C:Genetics:
 A:Gene: char
 A:Note: expressed in floral tissue
 C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 43.7%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 3.4;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 111 OKAIKMGQPKSKI 124

RESULT 11

STPUCA

naringenin-chalcone synthase (EC 2.3.1.74) A - garden petunia

N:Alternate names: chalcone synthase

C:Species: Petunia x hybrida (garden petunia)

C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000

C:Accession: J50308

R:Koebe, R.E.; Spelt, C.E.; van den Elzen, P.J.M.; Mol, J.N.M.

Gene 81, 245-257, 1989

A:Title: Cloning and molecular characterization of the chalcone synthase multigene family

A:Reference number: J50308; MUID:90034197; PMID:2806915

A:Accession: J50308

A:Molecule type: DNA

A:Residues: 1-389 <KOE>

A:Cross-references: GB:X14591; NID:920524; PIDN:CAA32731.1; PID:920525

A:Experimental source: strain Violet 30, leaf

C:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavonoid

C:Genetics:

A:Gene: chsA

A:Map position: V

A>Note: chsA is the major expressed member of the gene family in various floral tissues

C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 43.7%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 3.4;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 111 OKAIKMGQPKSKI 124

RESULT 12

JCS136

naringenin-chalcone synthase (EC 2.3.1.74) 2 - potato

N:Alternate names: chalcone synthase; CHS

C:Species: Solanum tuberosum (potato)

C>Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 05-May-2000

C:Accession: JCS136; PC4239

R:Jeon, J.H.; Kim, H.S.; Choi, K.H.; Jung, Y.H.; Jung, H.; Byun, S.M.

Biochem. Biotechnol. Biochem. 60, 1907-1910, 1996

A:Title: Cloning and characterization of one member of the chalcone synthase gene family

A:Reference number: JCS136; MUID:97141614; PMID:8987872

A:Accession: JCS136

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-389 <JCS1>

A:Cross-references: GB:U47378; NID:91470059; PIDN:AA05239.1; PID:91470060

A:Accession: PC4239

A:Status: preliminary

A:Molecule type: protein

A:Residues: 158-165;367-373 <JCS2>

C:Comment: This enzyme is important in the biosynthesis of all classes of flavonoids in C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A

Query Match 43.7%; Score 52; DB 2; Length 389;
Best Local Similarity 64.3%; Pred. No. 3.4;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 5 OKANKPMIOPKTKV 18

Db 111 OKAIKMGQPKSKI 124

RESULT 13

T07799

naringenin-chalcone synthase (EC 2.3.1.74) - common morning-glory

N:Alternate names: chalcone synthase

C:Species: Ipomoea purpurea (common morning-glory)

C>Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000

C:Accession: T07799

R:Fukada-Tanaka, S.; Hoshino, A.; Hisatomi, Y.; Habu, Y.; Haebe, M.; Iida, S.

Plant Cell Physiol. 38, 754-758, 1997

A:Title: Identification of new chalcone synthase genes for flower pigmentation in the Ja

A:Reference number: Z16140; MUID:97393496; PMID:9249990

A:Accession: T07799

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-389 <FUK>

A:Cross-references: EMBL:AB001827; NID:92329836; PIDN:BA421769.1; PID:92329837

A:Gene: CHD-B

C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 43.7%; Score 52; DB 2; Length 389;
Best Local Similarity 64.3%; Pred. No. 3.4;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 111 OKAIKMGQPKSKI 124

RESULT 14

SVSKCD

naringenin-chalcone synthase (EC 2.3.1.74) - garden snapdragon

N:Alternate names: chalcone synthase

C:Species: Antirrhinum majus (garden snapdragon)

C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000

C:Accession: S07312; A33217

R:Sommer, H.; Saedler, H.

Mol. Gen. Genet. 202, 429-434, 1986

A:Title: Structure of the chalcone synthase gene of Antirrhinum majus.

A:Reference number: S07312

A:Accession: S07312

A:Molecule type: DNA

A:Residues: 1-390 <SOM>

A:Cross-references: EMBL:X03710; NID:916015; PIDN:CAA27338.1; PID:916016

A:Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-390 <SOM2>

C:Genetics:

A:Gene: chs

A:introns: 60/1; 162/3

C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 43.7%; Score 52; DB 1; Length 390;
Best Local Similarity 64.3%; Pred. No. 3.4;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 111 OKAIKMGQPKSKI 124

RESULT 15

S39775

alpha-s2-caaein form a precursor - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 13-Aug-1999

C:Accession: S39775
R:Dawson, S.P.; Wilde, C.J.; Tighe, P.J.; Mayer, R.J.
Biochem. J. 296, 777-784, 1993
A:Title: Characterization of two novel casein transcripts in rabbit mammary gland.
A:Reference number: S39775; MUID:94107245; PMID:8280077
A:Accession: S39775
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-180 <DAW>
A:Cross-references: EMBL:X76907; NID:g439525; PIDN:CAA54228.1; PID:g439526
C:Superfamily: gamma-casein

Query Match 42.9%; Score 51; DB 2; Length 180;
Best Local Similarity 50.0%; Pred. No. 2.2;
Matches 10; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 VYOHKMKEMIQPKTKVIP 20
DB 99 VYOOQIVMNPWKHKVKTITYP 118

Search completed: July 30, 2003, 16:31:15
Job time : 14.375 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:09:50 ; Search time 41.375 Seconds
(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-2
Perfect score: 119
Sequence: 1 VYOHKAMKRWIQPKTKVLPY 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_19Jun03:*
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	21	21	AAV88338
2	119	100.0	24	15	AAV88339
3	119	100.0	24	21	AAV88339
4	119	100.0	25	13	AAV88340
5	119	100.0	25	13	AAV88340
6	119	100.0	31	18	AAV32219
7	119	100.0	31	23	AAV17466
8	119	100.0	222	18	AAV32220
9	119	100.0	222	23	AAV17468

10	108.5	91.2	26	20	AAV93885	Bifidobacterium bi
11	95	79.8	223	23	AAV17469	Alpha-S2 casein pr
12	95	79.8	223	23	AAV17470	Goat alpha-S2 case
13	95	79.8	223	23	AAV17471	Capra hircus alpha
14	95	79.8	223	23	AAV17472	Sheep alpha-S2 cas
15	95	79.8	223	23	AAV17475	Alpha-S2 casein pr
16	92	77.3	16	21	AAV88337	Internal peptide f
17	85	71.4	19	18	AAV32218	Internal peptide f
18	85	71.4	19	18	AAV17465	Alpha-S2 casein pr
19	59.5	50.0	234	23	AAV17474	Bovine alpha-S2 ca
20	59.5	49.6	178	23	AAV17472	Pig alpha-S2 casei
21	53	44.5	9	17	AAV04282	Rabbit alpha-S2 ca
22	52	43.7	360	22	AAV60169	Milk derived anti-
23	52	43.7	389	17	AAV96125	Petunia chalcone s
24	52	43.7	389	22	AAV74019	Chalcone synthase.
25	52	43.7	389	22	ABG73501	Cyclamen chalcone
26	50	42.0	291	22	AAV69092	Petunia chalcone s
27	49	41.2	8	17	AAV04278	Mouse GN protein s
28	49	41.2	13	23	AAV17464	Milk derived anti-
29	49	41.2	133	17	AAV04628	Bovine alpha-S2 ca
30	49	41.2	133	20	AAV80492	Vascular endotheli
31	49	41.2	133	20	AAV80492	Vascular endotheli
32	49	41.2	160	20	AAV86208	Human VEGF-B trunc
33	49	41.2	167	20	AAV86234	Human VEGF-B full
34	49	41.2	168	17	AAV04826	Heart vascular end
35	49	41.2	168	17	AAV00864	Heart VPR167. Ma
36	49	41.2	188	20	AAV86201	Human vascular end
37	49	41.2	188	20	AAV80490	Murine vascular en
38	49	41.2	188	23	AAV83404	Murine vascular en
39	49	41.2	195	17	AAV04827	Heart vascular end
40	49	41.2	195	20	AAV80491	Murine vascular en
41	49	41.2	195	20	AAV83405	Vascular endotheli
42	49	41.2	207	17	AAV04830	Vascular endotheli
43	49	41.2	207	17	AAV00863	Murine VPR186. Ma
44	49	41.2	207	20	AAV80494	Murine vascular en
45	49	41.2	207	22	AAV36296	Human vascular end

ALIGNMENTS

RESULT 1
AAV88338 standard; peptide; 21 AA.
AAV88338;
14-JUL-2000 (first entry)
Internal peptide fragment of bovine alpha-S2 casein #2.
Alpha-S2 casein; peptide production; biological fluid; milk; whey; blood;
antibacterial peptide; lactoferrin; antiviral; antitumour activity.
Bos sp.
WO200015655-A1.
23-MAR-2000.
15-SEP-1999; 99WO-EP07002.
15-SEP-1998; 98EP-0203107.
08-JUN-1999; 99EP-0201815.
(NIZO-) NIZO FOOD RES.
Visser S, Recio I;
WPI, 2000-271377/23.
Novel process for producing peptides with e.g. antimicrobial activity
from biological fluids such as milk, whey or blood comprises contacting

PT fluid with chromatographic medium to adsorb peptide domain of interest
 PS Claim 14, Page 22; 41pp; English.
 CC This sequence represents an internal fragment of bovine alpha-S2 caesin
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.

XX Sequence 21 AA;

Query Match 100.0%; Score 119; DB 21; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKAMKPMIQPKTKVIFY 21
 DB 1 VYOHQKAMKPMIQPKTKVIFY 21

RESULT 2
 AAR60481

ID AAR60481 standard; peptide; 24 AA.

XX AAR60481;

XX 19-APR-1995 (first entry)

XX Peptide used in tranquilliser.

XX Peptide; tranquilliser; food; foodstuff; mental stabilisation.

XX Synthetic.

XX JP06211689-A.

XX 02-AUG-1994.

XX 19-JAN-1993; 93JP-0024811.

XX 19-JAN-1993; 93JP-0024811.

XX (KANE) KANEBO LTD.

XX WPI; 1994-283376/35.

XX Synthetic peptide(s) used in tranquilliser - also used in the

XX production of a food for mental stabilisation.

XX Claim 3, Page 1; 5pp; Japanese.

XX The peptide may be used as a component of a tranquilliser which may

XX be administered orally and is low in toxicity. The peptide may also

XX stabilisation. See also AAR60480.

XX Sequence 24 AA;

Query Match 100.0%; Score 119; DB 15; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.7e-10;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKAMKPMIQPKTKVIFY 21
 DB 1 VYOHQKAMKPMIQPKTKVIFY 21

RESULT 3

ID AAY86339 standard; peptide; 24 AA.

XX AAY86339;

XX 14-JUL-2000 (first entry)

XX Internal peptide fragment of bovine alpha-S2 caesin #3.

XX Alpha-S2 caesin; peptide production; biological fluid; milk; whey; blood;
 XX antibacterial peptide; lactoferrin; antiviral; antitumour activity.

XX Bos sp.

XX WO200015655-A1.

XX 23-MAR-2000.

XX 15-SEP-1999; 99WO-EP07002.

XX 15-SEP-1998; 98EP-0203107.

XX 08-JUN-1999; 99EP-0201815.

XX (NIZO-) NIZO FOOD RES.

XX Visser S, Reclio I;

XX WPI; 2000-271377/23.

XX Novel process for producing peptides with e.g. antimicrobial activity
 XX from biological fluids such as milk, whey or blood comprises contacting
 XX fluid with chromatographic medium to adsorb peptide domain of interest

XX Claim 14; Page 22; 41pp; English.

XX This sequence represents an internal fragment of bovine alpha-S2 caesin

XX protein. The peptide is an example of a peptide with antibacterial

XX activity that can be produced by the process of the invention. The

XX invention relates to a process for producing peptides from biological

XX fluids. The process comprises chromatography of the biological fluid, in

XX situ hydrolysis of selectively bound peptides, washing to remove unbound

XX peptide, and elution of the peptides of interest. The process is used for

XX producing peptides from biological fluids, such as milk, whey or blood.

XX For example, the process can be used to produce antibacterial peptides

XX derived from lactoferrin, using cheese whey as a starting material. The

XX peptides obtained have preferably antimicrobial and/or antiviral and/or

XX antitumour activity. The process of the invention is relatively simple

XX and generally economically and technically more attractive than those

XX methods previously used. The method provides high yield peptides with a

XX selected activity of interest without the need for intermediate

XX purification of the precursor protein.

XX Sequence 24 AA;

Query Match 100.0%; Score 119; DB 21; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.7e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKAMKPMIQPKTKVIFY 21
 DB 1 VYOHQKAMKPMIQPKTKVIFY 21

RESULT 4

```

AAR20055
ID AAR20055 standard; peptide; 25 AA.
XX
AC AAR20055;
XX
DT 25-MAR-2003 (updated)
XX
DT 26-MAR-1992 (first entry)
XX
DE Casein peptide.
XX
KW Casein; platelet; aggregation; thrombosis; collagen.
XX
OS Synthetic.
XX
PN JP03255095-A.
XX
PD 13-NOV-1991.
XX
PE 02-MAR-1990; 90JP-0052554.
XX
PR 02-MAR-1990; 90JP-0052554.
XX
PA (KANEBO LTD.
XX
DR WPI; 1992-002669/01.
XX
PT Casein peptide(s) for treating thrombosis - as inhibitors of
PT platelet aggregation caused by adenosine-5'-diphosphate and
PT collagen and as biochemical reagents
XX
PS Claim 1; Page 1; BPP; Japanese.
XX
CC Leu25 may be omitted. The peptide and its salts have inhibiting
CC activity against platelet aggregation caused by adenosine-5'-
CC diphosphate and by collagen. They are useful for prevention and
CC treatment of thrombosis and are also useful as a biochemical reagent.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 25 AA;

Query Match 100.0%; Score 119; DB 13; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKAMKPMIOPKTKVIRY 21
DB 1 VYOHQKAMKPMIOPKTKVIRY 21

RESULT 5
AAY88340
ID AAY88340 standard; peptide; 25 AA.
XX
AC AAY88340;
XX
DT 14-JUL-2000 (first entry)
XX
DE C-terminal peptide fragment of bovine alpha-S2 casein.
XX
KW Alpha-S2 casein; peptide production; biological fluid; milk; whey; blood;
KW antibacterial peptide; lactoferrin; antiviral; antitumour activity.
XX
OS Bos sp.
XX
PN WO200015655-A1.
XX
PD 23-MAR-2000.
XX
PE 15-SEP-1999; 99MO-EP07002.
XX
PR 15-SEP-1998; 98EP-0203107.
XX
PR 08-JUN-1999; 99EP-0201815.
XX
PT

```

```

PA (NIZO-) NIZO FOOD RES.
XX
PI Visser S, Recio I.
XX
DR WPI; 2000-271377/23.
XX
PT Novel process for producing peptides with e.g. antimicrobial activity
PT from biological fluids such as milk, whey or blood comprises contacting
PT fluid with chromatographic medium to adsorb peptide domain of interest
XX
PS Claim 14; Page 22; 41pp; English.
XX
CC This sequence represents a C-terminal fragment of bovine alpha-S2 casein
CC protein. The peptide is an example of a peptide with antibacterial
CC activity that can be produced by the process of the invention. The
CC invention relates to a process for producing peptides from biological
CC fluids. The process comprises chromatography of the biological fluid, in
CC situ hydrolysis of selectively bound peptides, washing to remove unbound
CC peptide, and elution of the peptides of interest. The process is used for
CC producing peptides from biological fluids, such as milk, whey or blood.
CC For example, the process can be used to produce antibacterial peptides
CC derived from lactoferrin, using cheese whey as a starting material. The
CC peptides obtained have preferably antimicrobial and/or antiviral and/or
CC antitumour activity. The process of the invention is relatively simple
CC and generally economically and technically more attractive than those
CC methods previously used. The method provides high yield peptides with a
CC selected activity of interest without the need for intermediate
CC purification of the precursor protein.
XX
SQ Sequence 25 AA;

Query Match 100.0%; Score 119; DB 21; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKAMKPMIOPKTKVIRY 21
DB 1 VYOHQKAMKPMIOPKTKVIRY 21

```

```

RESULT 6
AAW32219
ID AAW32219 standard; peptide; 31 AA.
XX
AC AAW32219;
XX
DT 03-FEB-1998 (first entry)
XX
DE Alpha-S2 casein precursor C-terminal peptide fragment #5.
XX
KW Alpha-S2 casein precursor; growth promoting; mitogenic assay;
KW platelet-derived growth factor; insulin-like derived growth factor;
XX
OS Synthetic.
XX
OS Bos taurus.
XX
PN WO9716460-A1.
XX
PD 09-MAY-1997.
XX
PF 31-OCT-1996; 96MO-GB02658.
XX
PR 31-OCT-1995; 95GB-0022302.
XX
PA (UNIL-) UNIV LIVERPOOL.
XX
PI Liu Q, Smith JA, Wilkinson, MC;
XX
DR WPI; 1997-272048/24.
XX
PT Manufacture of medicament or foodstuff for promoting growth - using
PT peptide(s) with a sequence identical to the C-terminal end of an

```

PT alpha-S2 casein precursor
 XX
 PS Claim 12; Page 21; 33pp; English.
 CC
 CC The present sequence, which is substantially identical to the C-terminal
 CC end of an alpha-S2 casein precursor, was found after storage in PBS to
 CC exhibit growth promoting activity for rat mammary fibroblast cell line
 CC (Rama 27), which is not significantly stimulated by platelet-derived
 CC growth factor or insulin-like growth factor. The activity of the peptide
 CC increased when maintained at alkaline pH. By way of contrast, alpha-S2
 CC casein was inactive in a mitogenic assay. This peptide may be used in the
 CC manufacture of a medicament or foodstuff for promoting growth in
 CC humans or animals.
 XX
 SQ Sequence 31 AA;
 Query Match 100.0%; Score 119; DB 18; Length 31;
 Best Local Similarity 100.0%; Pred. No. 2,2e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKPMIQPKTKVPIY 21
 DB 7 VYQHOKAMKPMIQPKTKVPIY 27
 RESULT 7
 AAE17466
 ID AAE17466 standard; peptide; 31 AA.
 AC
 XX AAE17466;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Bovine alpha-S2 casein precursor protein C-terminal fragment #7.
 XX
 XX Bovine; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 XX collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 XX chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Bos sp.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin
 XX
 PS Claim 9; Page 22; 27pp; English.
 CC
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptide stimulates the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is a peptide
 CC derived from the C-terminal of bovine alpha-S2 casein precursor protein.

XX
 SQ Sequence 31 AA;
 Query Match 100.0%; Score 119; DB 23; Length 31;
 Best Local Similarity 100.0%; Pred. No. 2,2e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKPMIQPKTKVPIY 21
 DB 7 VYQHOKAMKPMIQPKTKVPIY 27
 RESULT 8
 AAW32220
 ID AAW32220 standard; protein; 222 AA.
 XX
 AC AAW32220;
 XX
 DT 03-FEB-1998 (first entry)
 XX
 DE Bovine alpha-S2 casein precursor.
 XX
 XX Bovine alpha-S2 casein precursor; growth promoting; mitogenic assay;
 XX platelet-derived growth factor; insulin-like derived growth factor;
 XX
 OS Bos taurus.
 XX
 PN WO9716460-A1.
 XX
 PD 09-MAY-1997.
 XX
 PF 31-OCT-1996; 96WO-GB02658.
 XX
 PR 31-OCT-1995; 95GB-0022302.
 XX
 PA (UNIT-) UNITV LIVERPOOL.
 XX
 PI Lin Q, Smith JA, Wilkinson MC;
 XX
 DR WPI; 1997-272048/24.
 XX
 PT Manufacture of medicament or foodstuff for promoting growth - using
 PT peptide(s) with a sequence identical to the C-terminal end of an
 PT alpha-S2 casein precursor
 XX
 PS Disclosure; Page 3; 33pp; English.
 XX
 CC The present sequence represents bovine alpha-S2 casein precursor.
 CC Peptides having an amino acid sequence which is substantially identical
 CC to the C-terminal end of an alpha-S2 casein precursor, are used for
 CC manufacture of a medicament or foodstuff for promoting growth in humans
 CC or animals.
 XX
 SQ Sequence 222 AA;
 Query Match 100.0%; Score 119; DB 18; Length 222;
 Best Local Similarity 100.0%; Pred. No. 1.8e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKPMIQPKTKVPIY 21
 DB 198 VYQHOKAMKPMIQPKTKVPIY 218
 RESULT 9
 AAE17468
 ID AAE17468 standard; protein; 222 AA.
 XX
 AC AAE17468;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Bovine alpha-S2 casein precursor protein.

XX Bovine; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KW chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Bos sp.
 XX WO200202133-A2.
 XX
 XX 10-JAN-2002.
 XX
 XX 13-JUN-2001; 2001WO-GB02601.
 XX
 XX 30-JUN-2000; 2000GB-0016189.
 XX
 XX (PEPS-) PEPSYN LTD.
 XX
 XX Smith JA;
 XX
 XX WPI; 2002-154690/20.
 XX
 XX Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 XX Claim 8; Page 6; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging.
 CC particularly wrinkling of the skin. The present sequence is bovine
 CC alpha-S2 casein precursor protein.
 CC
 CC Sequence 222 AA;
 SQ

Query Match 100.0%; Score 119; DB 23; Length 222;
 Best Local Similarity 100.0%; Pred. No. 1.8e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VYOHOKAMKPMIOPKTKVTPY 21
 |||||
 DB 198 VYOHOKAMKPMIOPKTKVTPY 218

RESULT 10
 AA93885
 ID AAW93885 standard; peptide; 26 AA.
 XX
 AC AAW93885;
 XX
 DT 25-JUN-1999 (first entry)
 XX
 DE Bifidobacterium bifidus stimulating peptide 21.
 XX
 KW Bifidogenic peptide; protease; treatment; microbe-related disease;
 KW bacteria; fungi; yeast; protozoa; virus; mycoplasma; filaria; plasmodia;
 KW infection; inflammation; microbial induced tumour; degenerative disorder;
 KW diarrhoea; colic; oral microflora; intestinal microflora; caries;
 KW vaginal microflora.
 XX
 OS Bifidobacterium bifidus.
 XX
 XX WO9914231-A2.
 XX
 XX 25-MAR-1999.
 XX
 XX

PF 16-SEP-1998; 98WO-EP05899.
 XX
 XX 11-FEB-1998; 98DE-1005385.
 PR 16-SEP-1997; 97DE-1040504.
 XX
 XX (FORS/) FORSMANN W.
 XX
 XX Foresmann W, Lapeke C, Zucht H;
 PI WPI; 1999-244022/20.
 XX
 XX
 XX Milk-derived peptides that stimulate Bifidobacterium bifidus
 PT
 PT Claim 2; Page 3; 25pp; German.
 PS
 XX
 CC This invention describes milk-derived bifidogenic peptides and their
 CC active derivatives or fragments, and combinations of them produced by
 CC chemical coupling. Such are produced from bovine or human milk by
 CC treatment for 2 hr with proteases, then centrifuging to remove fat and
 CC acidifying to pH 2 to precipitate proteins. The solution phase is then
 CC subjected to reverse-phase high-performance liquid chromatography (HPLC)
 CC and cation-exchange HPLC, the fractions adjusted to salt content below
 CC 25 mM (by dialysis or reverse-phase HPLC) and tested for activity by
 CC growing Bifidobacterium bifidus and Bacterioides coli in presence of the
 CC fractions. Those fractions for which (Bw-B0) - (Bw-E0) is at least 0.15
 CC are selected where Bw = germ count after 16 hr culture of B. bifidus in
 CC 50% Elliker broth containing peptide at 0.2 mg/ml, B0 = germ count under
 CC similar conditions in a peptide-free control, Bw = germ count after 16
 CC hr culture of E. coli in 9/1 tryptic broth containing peptide at
 CC 0.2 mg/ml, B0 = germ count under similar conditions in a peptide-free
 CC control. The peptides AAW93885-W93888 are used to treat microbe-related
 CC diseases caused by bacteria, fungi, yeast, protozoa, viruses,
 CC mycoplasma, filaria and plasmodia, e.g. infections, inflammation,
 CC microbially induced tumours or degenerative disorders, diarrhoea, colic,
 CC abnormalities in oral, intestinal or vaginal microflora, or caries.
 CC
 CC Sequence 26 AA;
 SQ

Query Match 91.2%; Score 108.5; DB 20; Length 26;
 Best Local Similarity 95.5%; Pred. No. 6.2e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 VYOHOKAM-KPMIOPKTKVTPY 21
 |||||
 DB 1 VYOHOKAMKPMIOPKTKVTPY 22

RESULT 11
 AA817469
 ID AA817469 standard; Protein; 223 AA.
 XX
 AC AA817469;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Alpha-S2 casein precursor (alpha-S2-CN) protein #1.
 XX
 KW Alpha-S2 casein; alpha-S2-CN; dermatological; antiinflammatory; cosmetic;
 KW fibroblast; collagen; keratinocyte; skin regeneration; medicament; aging;
 KW toothpaste; chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS unidentified.
 XX
 XX WO200202133-A2.
 XX
 XX 10-JAN-2002.
 XX
 XX 13-JUN-2001; 2001WO-GB02601.
 XX
 XX 30-JUN-2000; 2000GB-0016189.
 XX
 XX (PEPS-) PEPSYN LTD.
 XX
 XX

PI Smith JA;
 XX WPI; 2002-154690/20.
 XX
 XX Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Disclosure; Page 8; 27pp; English.
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is alpha-S2
 CC casein precursor (alpha-S2-CN) protein.
 CC
 SQ Sequence 223 AA;
 Query Match 79.8%; Score 95; DB 23; Length 223;
 Best Local Similarity 81.0%; Pred. No. 5.7e-06;
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 1 VYOHOKAMKPMWIOPTKVIPIY 21
 DB 199 VDOHOKAMKPMWIOPTKVIPIY 219
 RESULT 12
 AAE17470
 ID AAE17470 standard; Protein; 223 AA.
 XX
 AC AAE17470;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Goat alpha-S2 casein E precursor protein.
 XX
 KW Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KW chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Capra hircus.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Claim 8; Page 8; 27pp; English.
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2

CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat
 CC alpha-S2 casein E precursor protein.
 CC
 SQ Sequence 223 AA;
 Query Match 79.8%; Score 95; DB 23; Length 223;
 Best Local Similarity 81.0%; Pred. No. 5.7e-06;
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 1 VYOHOKAMKPMWIOPTKVIPIY 21
 DB 199 VDOHOKAMKPMWIOPTKVIPIY 219
 RESULT 13
 AAE17471
 ID AAE17471 standard; Protein; 223 AA.
 XX
 AC AAE17471;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Capra hircus alpha-S2 casein C precursor protein.
 XX
 KW Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KW chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Capra hircus.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Claim 8; Page 8; 27pp; English.
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat
 CC alpha-S2 casein C precursor protein.
 CC
 SQ Sequence 223 AA;
 Query Match 79.8%; Score 95; DB 23; Length 223;

Best Local Similarity 81.0%; Pred. No. 5.7e-06;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIQPKTKVLPY 21
DB 199 VDOHQAMKPMPTQPKTNALPY 219

RESULT 14

ID AAE17473 standard; Protein; 223 AA.

AC AAE17473;

DT 22-APR-2002 (first entry)

DE Sheep alpha-S2 casein precursor protein.

KW Sheep; alpha-S2 casein; dermatological; antiinflammatory; fibroblast; collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging; chewing gum; cosmetic; wrinkling; periodontal disease.

OS Ovis sp.

PN WO200202133-A2.

PD 10-JAN-2002.

PF 13-JUN-2001; 2001WO-GB02601.

PR 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD.

PI Smith JA;

DR WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in alpha-S2 casein precursor in the manufacture of a medicament for alleviating or preventing periodontal disease and an effect of aging in skin

PS Claim 8; Page 9; 27pp; English.

CC The invention relates to a composition comprising a peptide or its derivative. The peptide contains an amino acid sequence from alpha-S2 casein precursor. The peptides stimulate the growth of fibroblasts, and thus the synthesis and secretion of collagen. The peptides also stimulate the growth of keratinocytes, which aid in the formation and regeneration of skin surface. The peptide is useful in the manufacture of a medicament in the form of a toothpaste or a chewing gum, for alleviating or preventing periodontal disease and a medicament in the form of a cosmetic composition for alleviating or preventing an effect of aging, particularly wrinkling of the skin. The present sequence is sheep alpha-S2 casein precursor protein.

SQ Sequence 223 AA;

Query Match 79.8%; Score 95; DB 23; Length 223;
Best Local Similarity 81.0%; Pred. No. 5.7e-06;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIQPKTKVLPY 21
DB 199 VDOHQAMKPMPTQPKTNALPY 219

RESULT 15

ID AAE17475 standard; Protein; 223 AA.

AC AAE17475;

XX 22-APR-2002 (first entry)

DE Alpha-S2 casein precursor (alpha-S2-CN) protein #2.

KW Alpha-S2 casein; alpha-S2-CN; dermatological; antiinflammatory; cosmetic; fibroblast; collagen; keratinocyte; skin regeneration; medicament; aging; toothpaste; chewing gum; cosmetic; wrinkling; periodontal disease.

OS Unidentified.

PN WO200202133-A2.

PD 10-JAN-2002.

PF 13-JUN-2001; 2001WO-GB02601.

PR 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD.

PI Smith JA;

DR WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in alpha-S2 casein precursor in the manufacture of a medicament for alleviating or preventing periodontal disease and an effect of aging in skin

PS Disclosure; Page 9; 27pp; English.

CC The invention relates to a composition comprising a peptide or its derivative. The peptide contains an amino acid sequence from alpha-S2 casein precursor. The peptides stimulate the growth of fibroblasts, and thus the synthesis and secretion of collagen. The peptides also stimulate the growth of keratinocytes, which aid in the formation and regeneration of skin surface. The peptide is useful in the manufacture of a medicament in the form of a toothpaste or a chewing gum, for alleviating or preventing periodontal disease and a medicament in the form of a cosmetic composition for alleviating or preventing an effect of aging, particularly wrinkling of the skin. The present sequence is sheep alpha-S2 casein precursor (alpha-S2-CN) protein.

SQ Sequence 223 AA;

Query Match 79.8%; Score 95; DB 23; Length 223;
Best Local Similarity 81.0%; Pred. No. 5.7e-06;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIQPKTKVLPY 21
DB 199 VDOHQAMKPMPTQPKTNALPY 219

Search completed: July 30, 2003, 16:23:26
Job time : 41.375 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 40.7143 Seconds

(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-3

Sequence: 1 VYOHQKMKPWIOPTKTVIPYRVY 24

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

1: SPREMBL_23:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.rodent:*
13: sp.virus:*
14: sp.unclassified:*
15: sp.ivirus:*
16: sp.bacteriophage:*
17: sp.archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	88.1	210	6	062825 Bubalus bub
2	111	82.2	223	6	09MYU6 Q9MYU6
3	102	75.6	223	6	09MYU6 Q9MYU6
4	102	75.6	223	6	09MYU6 Q9MYU6
5	53	39.3	935	3	094623 capra hircu
6	52	38.5	80	10	081534 saccharum c
7	52	38.5	85	10	081399 oryzae c
8	52	38.5	340	10	096568 ipomoea pur
9	52	38.5	379	10	094065 perilla fru
10	52	38.5	366	10	023923 digitalis l
11	52	38.5	388	10	09F537 torenia hyb
12	52	38.5	389	10	08LPI9 nierenbergi
13	52	38.5	389	10	043040 petunia hyb
14	52	38.5	389	10	093XP8 nicotiana t
15	52	38.5	389	10	09MSB2 petunia hyb
16	52	38.5	390	10	09LUB2 scutellaria

17	52	38.5	390	10	048564 scutellaria
18	52	38.5	390	10	022122 scutellaria
19	51	37.8	394	10	093V86 humulus lup
20	51	37.8	399	10	094LW8 humulus lup
21	50	37.0	311	5	019257 caenorhabdi
22	50	37.0	633	11	09WV66 mus musculu
23	50	37.0	1877	3	09USR3 Oryza sativa
24	49.5	36.7	325	16	09K947 bacillus ha
25	49	36.3	322	5	044074 ascaris suu
26	49	36.3	342	16	08YU06 anabaena sp
27	48.5	35.9	286	10	08W2F2 arabidopsis
28	48	35.6	183	16	09CP51 pasteurilla
29	48	35.6	208	16	08F8G2 leprospira
30	48	35.6	225	4	09BTR9 Oribacterium
31	48	35.6	328	16	08CP60 staphylococ
32	48	35.6	374	10	081476 brassica na
33	48	35.6	392	10	093YX5 vitis sp. c
34	48	35.6	392	10	09PEK7 vitis labru
35	48	35.6	392	10	08LPP4 parthenocis
36	48	35.6	392	10	09PRW2 vitis ripar
37	48	35.6	392	10	09SPW2 vitis ripar
38	48	35.6	392	10	08LPP2 parthenocis
39	48	35.6	392	10	09S982 vitis. still
40	48	35.6	392	10	08LPP3 cisus rhom
41	48	35.6	392	10	094G58 vitis vinif
42	48	35.6	392	10	0944W8 vitis sp. c
43	48	35.6	392	10	0944W7 vitis sp. c
44	48	35.6	395	10	09FR70 allaria pe
45	48	35.6	408	10	09SLY0 pelilotum nu

ALIGNMENTS

RESULT 1	062825	PRELIMINARY;	PRT;	210 AA.
AC	062825;			
DT	01-AUG-1998 (TREMBLrel. 07, Created)			
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	AS2-casein (Fragment).			
OS	Bubalus bubalis (Domestic water buffalo).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bubalus.			
OX	NCBI_TaxID=89462;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISUB=Mammary gland;			
RA	Dae P., Jain S., Garg L.C.;			
RT	"Cloning and nucleotide sequence of cDNA encoding AS2-casein in B.			
RT	bubalis";			
RL	Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AJ005431; CAA06534.2; -			
DR	InterPro; IPR001588; Casein.			
DR	Pfam; PF00363; caseins; 2.			
FT	NON_TER			
SQ	SEQUENCE 210 AA; 24700 MW; 05DEF95963F1132C CRC64;			
Query Match	88.1%; Score 119; DB 6; Length 210;			
Best Local Similarity	87.5%; Pred. No. 2.4e-10;			
Matches	21; Conservative 1; Mismatches 2; Indels 0; Gaps 0;			
QY	1 VYOHQKMKPWIOPTKTVIPYRVY 24			
DB	186 VYOHQKMKPWIOPTKTVIPYRVY 209			
RESULT 2	09MYU6	PRELIMINARY;	PRT;	223 AA.
ID	09MYU6			
AC	09MYU6;			

```

DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Alpha s2-casein.
GN CSNIS2.
OC Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OK NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA Lagonigro R., Pilla F., Matasino D., Zullo A.;
RT "A new allele of goat alpha s2-casein gene.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ289716; CAB94236.1; -
DR EMBL; AJ289716; CAB94236.1; -
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN_ALPHA_BETA. 1.
SQ SEQUENCE 223 AA; 26403 MW; 0E1F83F24DA85E2 CRC64;

Query Match
Best Local Similarity 82.2%; Score 111; DB 6; Length 223;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPKTKVPIPVRY 24
DB 199 VDOHOKAMKPMIOPKTKVPIPVRY 222

RESULT 3
Q9TT07 PRELIMINARY; PRT; 223 AA.
AC O9TT07;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE Alpha s2-casein.
GN CSNIS2.
OC Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OK NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA Velti C.C., Pilla F., Lagonigro R.R.;
RT "A new allele of goat alpha s2-casein.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ249995; CAB59920.1; -
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN_ALPHA_BETA. 1.
SQ SEQUENCE 223 AA; 26433 MW; CE9F4DC8D768293 CRC64;

Query Match
Best Local Similarity 79.2%; Score 102; DB 6; Length 223;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPKTKVPIPVRY 24
DB 199 VDOHOKAMKPMIOPKTKVPIPVRY 222

RESULT 4
Q9GK07 PRELIMINARY; PRT; 223 AA.
AC Q9GK07;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Alpha s2-casein.
GN CSNIS2.

```

```

OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OK NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA Velti C., Pilla F., Lagonigro R.;
RT "A new allele of alpha s2-casein.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ287310; CAC21704.2; JOINED.
DR EMBL; AJ287311; CAC21704.2; JOINED.
DR EMBL; AJ242728; CAC21704.2; JOINED.
DR EMBL; AJ297312; CAC21704.2; JOINED.
DR EMBL; AJ297313; CAC21704.2; JOINED.
DR EMBL; AJ297313; CAC21704.2; JOINED.
DR EMBL; AJ242527; CAC21704.2; JOINED.
DR EMBL; AJ287315; CAC21704.2; JOINED.
DR EMBL; AJ287316; CAC21704.2; JOINED.
DR EMBL; AJ242526; CAC21704.2; JOINED.
DR EMBL; AJ242528; CAC21704.2; JOINED.
DR EMBL; AJ242533; CAC21704.2; JOINED.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN_ALPHA_BETA. 1.
SQ SEQUENCE 223 AA; 26432 MW; CE97658BD7688C9D CRC64;

Query Match
Best Local Similarity 75.6%; Score 102; DB 6; Length 223;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPKTKVPIPVRY 24
DB 199 VDOHOKAMKPMIOPKTKVPIPVRY 222

RESULT 5
O94623 PRELIMINARY; PRT; 935 AA.
AC O94623;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DE Hypothetical 106.5 kDa protein C1347.01C in chromosome II.
GN SPBC1347.01C OR SPBC215.16C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OK NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=972;
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Rieger M.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 731-935 FROM N.A.
RC STRAIN=972;
RA Lyne M., Rajandream M.A., Barrell B.G., Rieger M.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: TO C.ELEGANS ZK675.2.
CC -1- SIMILARITY: BELONGS TO THE UMG (DNA REPAIR) FAMILY.
CC -1- SIMILARITY: CONTAINS 1 BRCT DOMAIN.
DR EMBL; AL035548; CAB37432.1; -
DR EMBL; AL035548; CAB37432.1; -
DR HSP; P96022; 11M4.
DR GeneDB; SPombe; SPBC1347.01c; -
DR InterPro; IPR001357; BRCT.
DR InterPro; IPR001126; UMG_1like.
DR Pfam; PF00533; BRCT. 1.
DR Pfam; PF00817; IMS. 1.
DR SMART; SMO0292; BRCT. 1.

```

DR PROSITE: PS50172; BRCT. 1.
 DR PROSITE: PS50173; UROC. 1.
 KW Hypothetical protein; Nuclear protein.
 FT DOMAIN 59 147 BRCT.
 FT DOMAIN 279 460 UROC.
 SQ SEQUENCE 935 AA; 106525 MW; B1A472DDA37085FC CRC64;

Query Match 39.3%; Score 53; DB 3; Length 935;
 Best Local Similarity 37.0%; Pred. No. 16;
 Matches 10; Conservative 8; Mismatches 5; Indels 4; Gaps 2;

QY 2 YOHOKAMP-WIQ--PKTVIPYRY 24
 DB 119 WKHOKVVKPEMIVDCIKKKILPWINY 145

RESULT 6

ID 081534 PRELIMINARY; PRT; 80 AA.
 AC 081534;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Chalcone synthase (Fragment).
 GN CHS.
 OS Saccharum officinarum (Sugarcane).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACAD clade; Panicoideae; Andropogoneae; Saccharum.
 OC NCBI_TaxID=4547;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veronessi C.; Thalanam P.;
 RT "Sugarcane resistance to Striga hermorrhica.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBD databases.
 DR EMBL; AF079174; AAC27797.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C synthase.
 DR Pfam; PF00195; Chal_st11_synC; 1.
 DR ProDom; PD000453; N-C synthase; 1.
 FT NON_TER 1 80
 FT NON_TER 80
 SQ SEQUENCE 80 AA; 8744 MW; 7B94776D2A46F418 CRC64;

Query Match 38.5%; Score 52; DB 10; Length 80;
 Best Local Similarity 64.3%; Pred. No. 1.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QXAKMPWIOPKTV 18
 DB 33 QXAKEMWQPKSKI 46

RESULT 7

ID 081399 PRELIMINARY; PRT; 85 AA.
 AC 081399;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Chalcone synthase (fragment).
 GN CHS.
 OS Orobanche cumana.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Orobanchaceae; Orobancheae; Orobanche.
 OC NCBI_TaxID=78542;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veronessi C.; Labrousse P.; Thalanam P.;
 RT "Histological, physiological and molecular aspects of sunflower
 resistance to Orobanche cumana.";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBD databases.

DR EMBL; AF074401; AAC26132.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C synthase.
 DR Pfam; PF00195; Chal_st11_synC; 1.
 DR ProDom; PD000453; N-C synthase; 1.
 FT NON_TER 1 85
 FT NON_TER 85
 SQ SEQUENCE 85 AA; 9483 MW; B9A6AHEDD6D4777E CRC64;

Query Match 38.5%; Score 52; DB 10; Length 85;
 Best Local Similarity 64.3%; Pred. No. 1.9;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QXAKMPWIOPKTV 18
 DB 33 QXAKEMWQPKSKI 46

RESULT 8

ID Q96568 PRELIMINARY; PRT; 340 AA.
 AC Q96568;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Chalcone synthase.
 GN CHS-FIL.
 OS Ipomoea purpurea (Common morning glory).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Convolvulaceae; Ipomoea.
 OC NCBI_TaxID=4121;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISUB=flower buds;
 RA Rauber M.D.; Tiffin P.L.; Miller R.E.;
 RT "Regulation of anthocyanin gene expression in Ipomoea purpurea.";
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBD databases.
 DR EMBL; U74082; AAB41103.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C synthase.
 DR Pfam; PF00195; Chal_st11_synC; 1.
 DR Pfam; PF02797; Chal_st11_synC; 1.
 DR ProDom; PD000453; N-C synthase; 1.
 DR PROSITE; PS00441; CHALCONC SYNTH; 1.
 SQ SEQUENCE 340 AA; 37519 MW; 06682E6873DCFA39 CRC64;

Query Match 38.5%; Score 52; DB 10; Length 340;
 Best Local Similarity 64.3%; Pred. No. 8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QXAKMPWIOPKTV 18
 DB 111 QXAKEMWQPKSKI 124

RESULT 9

ID 004065 PRELIMINARY; PRT; 379 AA.
 AC 004065;
 DT 01-JUL-1997 (TREMBlrel. 04, Created)
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Chalcone synthase.
 GN Petilla frutescens.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Lamiales; Lamiaceae; Nepetoideae; Elsholtziaceae;
 OC Petilla.
 OC NCBI_TaxID=48386;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96088016; Pubmed=9426610;

RA Gong Z., Yamazaki M., Sugiyama M., Tanaka Y., Saito K.;
 RT "Cloning and molecular analysis of structural genes involved in
 RT anthocyanin biosynthesis and expressed in a forma-specific manner in
 RT *Pentilla frutescens*.";
 RL Plant Mol. Biol. 35:915-927(1997).
 DR EMBL; AB002582; BAA19548.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_still_synth; 1.
 DR Pfam; PF02797; Chal_still_synthC; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 379 AA; 4180 MW; 5E149C5530BD106 CRC64;
 QY Query Match 38.5%; Score 52; DB 10; Length 379;
 Best Local Similarity 64.3%; Pred. No. 8.9;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Db 5 OKAMKPMIOPKTV 18
 111 OKAIKMGOPKSKI 124
 RESULT 10
 023923 PRELIMINARY; PRT; 386 AA.
 ID 023923;
 AC 023923;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Chalcone synthase (Foxglove).
 OS Digitalis lanata (Foxglove).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Lamiales; Astringinaceae; Digitalis.
 NCBI_TaxId=49450;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=leaf;
 RA Theoringer C., Lindemann P., Luckner M.;
 RT "Expression of phenylalanine ammonia lyase, chalcone synthase and
 RT phytochrome in somatic embryogenesis of digitalis lanata.";
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ002526; CRA05512.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_still_synth; 1.
 DR Pfam; PF02797; Chal_still_synthC; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 DR NON_TER 1
 FT SEQUENCE 386 AA; 42190 MW; 0A88B8F8EDA82F8F CRC64;
 QY Query Match 38.5%; Score 52; DB 10; Length 386;
 Best Local Similarity 64.3%; Pred. No. 9.1;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Db 5 OKAMKPMIOPKTV 18
 106 OKAIKMGOPKSKI 119
 RESULT 11
 09FS37 PRELIMINARY; PRT; 388 AA.
 ID 09FS37;
 AC 09FS37;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Chalcone synthase.
 OS Torenia hybrida.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; lamids; Lamiales; Lamiales incertae sedis; Toreniaceae;
 OC Torenia.
 NCBI_TaxId=75807;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Summerwave Blue; TISSUE=Petal;
 RA Suzuki K., Xue H., Tanaka Y., Fukui Y., Fukushima-Mizutani M.,
 RA Murakami Y., Katsumoto Y., Teuda S., Kusumi T.;
 RT "Flower color modifications of Torenia hybrida by cosuppression of
 RT anthocyanin biosynthesis genes.";
 RL Mol. Breed. 6:239-246(2000).
 DR EMBL; AB012923; BAB20074.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_still_synth; 1.
 DR Pfam; PF02797; Chal_still_synthC; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 388 AA; 42402 MW; BD3A7B86C44A8B0F CRC64;
 QY Query Match 38.5%; Score 52; DB 10; Length 388;
 Best Local Similarity 64.3%; Pred. No. 9.2;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Db 5 OKAMKPMIOPKTV 18
 111 OKAIKMGOPKSKI 124
 RESULT 12
 08LP19 PRELIMINARY; PRT; 389 AA.
 ID 08LP19;
 AC 08LP19;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Chalcone synthase.
 GN CHS.
 OS Nierembergia sp. NB17.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Nierembergia.
 NCBI_TaxId=184877;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NB17;
 RA Ueyama Y., Katsumoto Y., Fukui Y., Ohkawa H., Kusumi T., Tanaka Y.;
 RT "Flower color modification of Nierembergia sp. by engineering
 RT flavonoid biosynthetic pathway.";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB078515; BAC10998.1; -.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_still_synth; 1.
 DR Pfam; PF02797; Chal_still_synthC; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 389 AA; 42616 MW; 62DA0CTFA07729E7 CRC64;
 QY Query Match 38.5%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 9.2;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Db 5 OKAMKPMIOPKTV 18
 111 OKAIKMGOPKSKI 124
 RESULT 13
 043040 PRELIMINARY; PRT; 389 AA.
 ID 043040;
 AC 043040;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Chalcone synthase.
 GN CHS-A.
 OS Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Solanaceae; Petunia.
 OX NCBI_TaxID=4102;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96156377; PubMed=8562849;
 RA Shao L., Li Y., Pan A., Cheng Z., Chen M.;
 RT "Molecular cloning, sequencing, and expression in *Escherichia coli* of
 the chalcone synthase gene."
 RL Chin. J. Biotechnol. 11:131-135(1995).
 DR EMBL; S80857; AAB36038.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro: IPR001099; N-C synthase.
 DR Pfam; PF02797; Chal_stil_synth; 1.
 DR Pfam; PF02797; Chal_stil_synth; 1.
 DR ProDom; PD000453; N-C synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 389 AA; 42642 MW; A931CF5B255A0A20 CRC64;

Query Match 38.5%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 9.2;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 OKANKPWIOPKTKV 18
 DB 111 OKAIKMGQPKSKI 124

RESULT 14
 O93XP8 PRELIMINARY; PRT; 389 AA.
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Chalcone synthase.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Flower;
 RA Hu H., Sung H., Su J.;
 RT "Cloning and Expression of a Tobacco Chalcone Synthase Gene."
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBD databases.
 DR EMBL; AF311783; AAK49457.1; -.
 DR InterPro: IPR001099; N-C synthase.
 DR Pfam; PF00195; Chal_stil_synth; 1.
 DR Pfam; PF02797; Chal_stil_synth; 1.
 DR ProDom; PD000453; N-C synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 389 AA; 42563 MW; 96C3D426786BD4 CRC64;

Query Match 38.5%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 9.2;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 OKANKPWIOPKTKV 18
 DB 111 OKAIKMGQPKSKI 124

RESULT 15
 O9MSB2 PRELIMINARY; PRT; 389 AA.
 AC Q9MSB2;
 AC Q9MSB2;

DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Chalcone synthase (EC 2.3.1.74).
 GN CHS.
 OS Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Solanaceae; Petunia.
 OX NCBI_TaxID=4102;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=violate; TISSUE=Corolla;
 RA Hsu Y.H., Durdan S.F., To K.Y.;
 RT "Cloning of a pigment-related cDNA encoding chalcone synthase,
 the chalcone isomerase, dihydroflavonol-4-reductase and cytochrome b5 from
 Petunia hybrida."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBD databases.
 DR EMBL; AF233638; AAF60297.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro: IPR001099; N-C synthase.
 DR Pfam; PF00195; Chal_stil_synth; 1.
 DR Pfam; PF02797; Chal_stil_synth; 1.
 DR ProDom; PD000453; N-C synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 DR Acyltransferase; Transferase.
 KW KW
 SQ SEQUENCE 389 AA; 42582 MW; 66135700259D7F0C CRC64;

Query Match 38.5%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 9.2;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 OKANKPWIOPKTKV 18
 DB 111 OKAIKMGQPKSKI 124

Search completed: July 30, 2003, 16:29:22
 Job time : 40.7143 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:10:10 ; Search time 7.85714 Seconds

(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-3

Sequence: 1 VYHQKAKMKWIOPTKTKIPYVRY 24

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_41.1

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	135	100.0	222	1	CAS2_BOVIN
2	111	82.2	223	1	CAS2_CAPRI
3	111	82.2	223	1	CAS2_SHEEP
4	70.5	52.2	235	1	CAS2_PIG
5	67	49.6	182	1	CAS2_RABBIT
6	55	40.7	193	1	CAS2_CAMDRA
7	52	38.5	389	1	CHS1_LYCIES
8	52	38.5	389	1	CHS2_LYCIES
9	52	38.5	389	1	CHS2_SOLTU
10	52	38.5	389	1	CHS2_PETHY
11	52	38.5	389	1	CHS2_SOLTU
12	52	38.5	389	1	CHS2_SOLTU
13	52	38.5	389	1	CHS2_SOLTU
14	52	38.5	389	1	CHS2_SOLTU
15	52	38.5	389	1	CHS2_SOLTU
16	52	38.5	389	1	CHS2_SOLTU
17	52	38.5	389	1	CHS2_SOLTU
18	52	38.5	389	1	CHS2_SOLTU
19	52	38.5	389	1	CHS2_SOLTU
20	52	38.5	389	1	CHS2_SOLTU
21	52	38.5	389	1	CHS2_SOLTU
22	52	38.5	389	1	CHS2_SOLTU
23	52	38.5	389	1	CHS2_SOLTU
24	52	38.5	389	1	CHS2_SOLTU
25	52	38.5	389	1	CHS2_SOLTU
26	52	38.5	389	1	CHS2_SOLTU
27	52	38.5	389	1	CHS2_SOLTU
28	52	38.5	389	1	CHS2_SOLTU
29	52	38.5	389	1	CHS2_SOLTU
30	52	38.5	389	1	CHS2_SOLTU
31	52	38.5	389	1	CHS2_SOLTU
32	52	38.5	389	1	CHS2_SOLTU
33	52	38.5	389	1	CHS2_SOLTU

34	48	35.6	392	1	THS2_VITVI
35	48	35.6	5171	1	PRP2_HUMAN
36	47	34.8	285	1	CHS2_MEDSA
37	47	34.8	311	1	CHS4_TRTSU
38	47	34.8	368	1	CHS1_SOYBN
39	47	34.8	368	1	CHS2_SOYBN
40	47	34.8	368	1	CHS3_SOYBN
41	47	34.8	368	1	CHS4_SOYBN
42	47	34.8	368	1	CHS5_SOYBN
43	47	34.8	368	1	CHS6_SOYBN
44	47	34.8	368	1	CHS7_SOYBN
45	47	34.8	368	1	CHS8_SOYBN

ALIGNMENTS

RESULT 1
CAS2_BOVIN STANDARD; PRT; 222 AA.
AC P02653; Q9TR51;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alpha-S2 casein precursor [contains: Casocidin-I].
GN CSN152.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=88188989; PubMed=2833669;
RA Stewart A.F., Bomsing U., Beattie C.W., Shah F., Willis I.M.,
RA Mackinnon A.G.;
RT "Complete nucleotide sequences of bovine alpha S2- and beta-casein
RT cDNAs: comparisons with related sequences in other species";
RM Mol. Biol. Evol. 4:231-241 (1987).
RN [2]
RP SEQUENCE OF 16-222 (A ALLELE).
RC TISSUE=Milk;
RX MEDLINE=77185633; PubMed=862906;
RA Britton G., Ribadeau-Dumas B., Mercier J.-C., Pellissier J.-P.,
RA Das B.C.;
RT "Complete amino acid sequence of bovine alphaS2-casein";
RM FEBS Lett. 76:274-279 (1977).
RN [3]
RP PARTIAL SEQUENCE (D ALLELE).
RC TISSUE=Milk;
RX MEDLINE=79239837; PubMed=469044;
RA Grosclaude F., Joudrier P., Mahe M.-F.;
RT "A genetic and biochemical analysis of a polymorphism of bovine alpha
RT S2-casein";
RM J. Dairy Res. 46:211-213 (1979).
RN [4]
RP SEQUENCE OF 165-203, AND CHARACTERIZATION OF CASOCIDIN.
RC TISSUE=Milk;
RX MEDLINE=9600204; PubMed=755666;
RA Zucht H.-D., Rada M., Adernann K., Meagert H.-U., Forsemann W.-G.;
RT "Casocidin-I: a casein-alpha s2 derived peptide exhibits antibacterial
RT activity";
RM FEBS Lett. 372:185-188 (1995).
RN [5]
RP FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
RP CALCIUM PHOSPHATE.
CC - FUNCTION: CASOCIDIN-I INHIBITS THE GROWTH OF ESCHERICHIA COLI AND
CC STREPTOCOCCUS CAROSUS.
CC - SUBCELLULAR LOCATION: Extracellular.
CC - TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC - MASS SPECTROMETRY: MW=4870; METHOD=Electrospray; RANGE=165-203.
CC - POLYMERISM: AT LEAST TWO ALLELES EXIST. THE SEQUENCE OF THE A
CC ALLELE IS SHOWN HERE. THE D ALLELE SEQUENCE DIFFERS FROM THAT
CC SHOWN IN HAVING A DELETION OF NINE RESIDUES, WHICH MAY BE 49-58,

CC 50-59, OR 51-60.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC -1- DATABASE: NAME=Protein Spotlight;
 CC NOTE=Issue 16 of November 2001;
 CC WWW="http://www.expasy.org/spotlight/articles/spot16016.html".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M16644; AAA0479.1; -
 CC PIR; J02008; KABOS2.
 CC InterPro; IPR001588; Casein.
 CC Pfam; PF00363; caseins; 2.
 CC PROSITE; PS00306; CASEIN ALPHA_BETA; 1.
 CC Milk; Phosphorylation; Signal; Repeat; Antibiotic.
 CC
 CC SIGNAL 1 15
 CC CHAIN 16 222 ALPHA-S2 CASEIN.
 CC PEPTIDE 165 203 CASOCIDIN-1.
 CC REPEAT 158 222
 CC MOD_RES 23 23 PHOSPHORYLATION.
 CC MOD_RES 24 24 PHOSPHORYLATION.
 CC MOD_RES 25 25 PHOSPHORYLATION.
 CC MOD_RES 31 31 PHOSPHORYLATION.
 CC MOD_RES 71 71 PHOSPHORYLATION.
 CC MOD_RES 72 72 PHOSPHORYLATION.
 CC MOD_RES 73 73 PHOSPHORYLATION.
 CC MOD_RES 76 76 PHOSPHORYLATION.
 CC MOD_RES 144 144 PHOSPHORYLATION.
 CC MOD_RES 146 146 PHOSPHORYLATION.
 CC CONFLICT 102 102 Q -> E (IN REF. 2).
 CC SEQUENCE 222 AA; 26019 MW; 81E7408AF1C12F7C CRC64;
 CC
 CC Query Match 100.0%; Score 135; DB 1; Length 222;
 CC Best Local Similarity 100.0%; Pred. No. 9, 6e-13;
 CC Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC 1 VYOHQAKMKPWIOPTKVIPIYVY 24
 CC 198 VYOHQAKMKPWIOPTKVIPIYVY 221
 CC
 CC RESULT 2
 CC CAS2_CAPHI STANDARD; PRT; 223 AA.
 CC ID CAS2_CAPHI
 CC AC P33049;
 CC DT 01-OCT-1993 (Rel. 27, Created)
 CC DT 01-OCT-1993 (Rel. 27, Last sequence update)
 CC DE 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Alpha-S2 casein precursor (Alpha-S2-CN).
 CC GN CSNIS2.
 CC OS Capra hircus (Goat).
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Caprinae; Capra.
 CC NCBI_TaxID=9925;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RA Boudiol C.; Bigsion G.; Mahe M.-F.; Prinz C.;
 CC "Characterization of goat allelic alpha-s2-caseins A and B: further
 CC evidence of the phosphorylation code of caseins.";
 CC Protein Seq. Data Anal. 5:213-218 (1993).
 CC RT
 CC RN

RN [3]
 RP SEQUENCE FROM N.A. (VARIANT C).
 RX MEDLINE=9503056; PubMed=794951;
 RA Boudiol C.; Bigsion G.; Mahe M.-F.; Prinz C.;
 RT "Biochemical and genetic analysis of variant C of caprine alpha s2-
 RL casein (Capra hircus).";
 RL Ann. Genet. 25:173-177 (1994).
 CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
 CC CALCIUM PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
 CC -1- POLYMORPHISM: THREE ALLELES OF ALPHA-S2 CASEIN ARE KNOWN: A, B AND
 CC C. THE FREQUENCIES OF THE ALLELES IS ESTIMATED TO BE 0.85, 0.04
 CC AND 0.11 IN THE FRENCH DAIRY BREEDS 'ALPINE' AND 'SAANEY'.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X65160; CAA46278.1; -
 CC DR EMBL; S74171; AAB32166.1; -
 CC PIR; I46985; I46995.
 CC PIR; S33881; UN0547.
 CC InterPro; IPR001588; Casein.
 CC Pfam; PF00363; caseins; 2.
 CC PROSITE; PS00306; CASEIN ALPHA_BETA; 1.
 CC Milk; Phosphorylation; Signal; Repeat; Polymorphism.
 CC
 CC SIGNAL 1 15
 CC CHAIN 16 223 ALPHA-S2 CASEIN.
 CC REPEAT 77 141
 CC REPEAT 159 223
 CC MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 77 77 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 145 145 PHOSPHORYLATION (BY SIMILARITY).
 CC MOD_RES 147 147 PHOSPHORYLATION (BY SIMILARITY).
 CC VARIANT 79 79 E -> K (IN VARIANT B).
 CC VARIANT 182 182 K -> I (IN VARIANT C).
 CC SEQUENCE 223 AA; 26389 MW; 187DEFA42FD688291 CRC64;
 CC
 CC Query Match 82.2%; Score 111; DB 1; Length 223;
 CC Best Local Similarity 83.3%; Pred. No. 3, 2e-09;
 CC Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 CC
 CC 1 VYOHQAKMKPWIOPTKVIPIYVY 24
 CC 199 VYOHQAKMKPWIOPTKVIPIYVY 222
 CC
 CC RESULT 3
 CC CAS2_SHEEP STANDARD; PRT; 223 AA.
 CC ID CAS2_SHEEP
 CC AC P04654;
 CC DT 13-AUG-1987 (Rel. 05, Created)
 CC DT 13-AUG-1987 (Rel. 05, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Alpha-S2 casein precursor.
 CC GN CSNIS2.
 CC OS Ovis aries (sheep).
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Caprinae; Ovis.
 CC NCBI_TaxID=9940;
 CC RN [1]


```

RP SEQUENCE FROM N.A.
RX MEDLINE=86104467; PubMed=3002499;
RA Boienard M., Petrisant G.;
RT "Complete sequence of ovine alpha s2-casein messenger RNA.";
RL Biochimie 67:1043-1051(1985).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03238; CA26983.1; -.
DR PIR; A25070; KASHS2.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN ALPHA BETA; 1.
DR MILK; Phosphorylation; Repeat; Signal.
FT SIGNAL 1 15
FT CHAIN 16 223 ALPHA-S2 CASEIN.
FT REPEAT 77 141
FT REPEAT 159 223
FT MOD RES 23 23 PHOSPHORYLATION (POTENTIAL).
FT MOD RES 24 24 PHOSPHORYLATION (POTENTIAL).
FT MOD RES 25 25 PHOSPHORYLATION (POTENTIAL).
FT MOD RES 72 72 PHOSPHORYLATION (POTENTIAL).
FT MOD RES 73 73 PHOSPHORYLATION (POTENTIAL).
FT MOD RES 74 74 PHOSPHORYLATION (POTENTIAL).
FT VARIANT 64 64 D -> N.
SQ SEQUENCE 223 AA; 26332 MW; 67212935E27426D7 CRC64;

Query Match 82.2%; Score 111; DB 1; Length 223;
Best Local Similarity 83.3%; Pred. No. 3.2e-09;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKMKPWIQPKTVIPYRY 24
Db 199 VDOHQKMKPWIQPKTVIPYRY 222

RESULT 4
ID CAS2_PIG STANDARD; PRT; 235 AA.
AC P39036.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-S2 casein precursor.
OS CSNIS2.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9623;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Mammary gland;
RX MEDLINE=92367960; PubMed=1503276;
RA Alexander L.J., Das Gupta N.A., Beattie C.W.;
RT "The sequence of porcine alpha s2-casein cDNA.";
RL Anim. Genet. 23:365-367(1992).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54975; CA38719.1; -.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN ALPHA BETA; 1.
DR MILK; Phosphorylation; Signal.
FT SIGNAL 1 15
FT CHAIN 16 235 ALPHA-S2 CASEIN.
FT MOD RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
FT MOD RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
FT MOD RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
FT MOD RES 32 32 PHOSPHORYLATION (BY SIMILARITY).
FT MOD RES 70 70 PHOSPHORYLATION (BY SIMILARITY).
FT MOD RES 71 71 PHOSPHORYLATION (BY SIMILARITY).
FT MOD RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
FT MOD RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
FT MOD RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 235 AA; 27570 MW; C903B760D184C14C CRC64;

Query Match 52.2%; Score 70.5; DB 1; Length 235;
Best Local Similarity 55.6%; Pred. No. 0.003;
Matches 15; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

QY 1 VYOHQKMKPWIQPKTVIPYRY 24
Db 208 VYOHQKMKPWIQPKTVIPYRY 234

RESULT 5
ID CAS3_RABIT STANDARD; PRT; 182 AA.
AC P50419.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Alpha-S2B casein precursor.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=New Zealand white; TISSUE=Mammary gland;
RX MEDLINE=94107245; PubMed=8280077;
RA Dawson S.P., Wilde C.J., Tighe P.J., Mayer R.J.;
RT "Characterization of two novel casein transcripts in rabbit mammary
RT gland.";
RL Biochem. J. 296:777-784(1993).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76909; CA54231.1; -.
DR PIR; S39776; S39776.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.

```


RA O'Neill S.D., Tong Y., Speerlein B., Forkmann G., Yoder J.I.;
 RT "Molecular genetic analysis of chalcone synthase in *Lycopersicon*
 RL esculentum and an anthocyanin-deficient mutant.",
 CC Mol. Gen. Genet. 224:279-288(1990).
 CC -1- FUNCTION: The primary product of this enzyme is 4',2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X55195; CAA38981.1; -;
 CC InterPro: IPR001099; N-C synthase.
 CC Pfam: PF00195; Chal_stil_synth; 1.
 CC DR Pfam: PF02797; Chal_stil_synth; 1.
 CC DR Pfam: PF000453; N-C_synthase; 1.
 CC DR PROSITE: PS00441; CHALCONE_SYNTH; 1.
 CC KM Flavonoid biosynthesis; Transferase; Acyltransferase;
 CC MultiGene family.
 CC FT ACT SITE 164 164
 CC FT ACT SITE 164 164 BY SIMILARITY.
 CC SQ SEQUENCE 389 AA; 42730 MW; F92E46B35FC32F CRC64;
 CC
 CC Query Match 38.5%; Score 52; DB 1; Length 389;
 CC Best Local Similarity 64.3%; Pred. No. 2.6;
 CC Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC QY 5 QKANKPMIOPKTV 18
 CC DB 111 QKAIKMGQPKSKI 124
 CC
 CC RESULT 9
 CC CHS2_SOLTU STANDARD; PRT; 389 AA.
 CC AC Q43188;
 CC DT 15-JUL-1999 (Rel. 38, Created)
 CC DT 15-JUL-1999 (Rel. 38, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2).
 CC GN CHS2.
 CC OS Solanum tuberosum (Potato).
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
 CC OC NCBI_TaxID=4113;
 CC OX [1]
 CC RN SEQUENCE FROM N.A.
 CC RP STRAIN=cv. Red Fontaine;
 CC RC MEDLINE=97141614; PubMed=9887872;
 CC RA Jeon J.-H., Kim H.S., Choi K.H., Joung Y.H., Joung H., Byun S.-M.;
 CC RT "Cloning and characterization of one member of the chalcone synthase
 CC gene family from *Solanum tuberosum* L.",
 CC RL BioSci. Biotechnol. Biochem. 60:1907-1910(1996).
 CC CC -1- FUNCTION: The primary product of this enzyme is 4',2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY

CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U47738; AAB05239.1; -;
 CC PIR: JCS136; JCS136.
 CC InterPro: IPR001099; N-C synthase.
 CC Pfam: PF00195; Chal_stil_synth; 1.
 CC DR Pfam: PF02797; Chal_stil_synth; 1.
 CC DR PROSITE: PS000453; N-C_synthase; 1.
 CC DR PROSITE: PS00441; CHALCONE_SYNTH; 1.
 CC KM Flavonoid biosynthesis; Transferase; Acyltransferase;
 CC MultiGene family.
 CC FT ACT SITE 164 164
 CC FT ACT SITE 164 164 BY SIMILARITY.
 CC SQ SEQUENCE 389 AA; 42476 MW; 41618F944958603 CRC64;
 CC
 CC Query Match 38.5%; Score 52; DB 1; Length 389;
 CC Best Local Similarity 64.3%; Pred. No. 2.6;
 CC Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC QY 5 QKANKPMIOPKTV 18
 CC DB 111 QKAIKMGQPKSKI 124
 CC
 CC RESULT 10
 CC CHS4_PETTY STANDARD; PRT; 389 AA.
 CC AC P08894;
 CC DT 01-NOV-1998 (Rel. 09, Created)
 CC DT 01-NOV-1998 (Rel. 09, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Chalcone synthase A (EC 2.3.1.74) (Naringenin-chalcone synthase A).
 CC GN CHS4.
 CC OS Petunia hybrida (Petunia).
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC OC Asteridae; lamids; Solanales; Solanaceae; Petunia.
 CC OC NCBI_TaxID=4102;
 CC OX [1]
 CC RN SEQUENCE FROM N.A.
 CC RP STRAIN=cv. Violet 30;
 CC RC MEDLINE=86286540; PubMed=3016642;
 CC RA Koes R.E., Spelt C.E., Reif H.U., van den Elzen P.J.M., Veltkamp E.,
 CC RT "Floral tissue of *Petunia hybrida* (V30) expresses only one member of
 CC the chalcone synthase multigene family.",
 CC RL Nucleic Acids Res. 14:5229-5239(1986).
 CC RN [12]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=cv. Violet 30; TISSUE=leaf;
 CC RC MEDLINE=90034197; PubMed=2806915;
 CC RA Koes R.E., Spelt C.E., van den Elzen P.J.M., Mol J.N.M.;
 CC RT "Cloning and molecular characterization of the chalcone synthase
 CC RT multigene family of *Petunia hybrida*.",
 CC RL Gene 81:245-257(1989).
 CC CC -1- FUNCTION: The primary product of this enzyme is 4',2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- TISSUE SPECIFICITY: MAJOR EXPRESSED MEMBER OF THE GENE FAMILY IN

ID	CHSE_IPOPU	STANDARD;	PRT;	389 AA.
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-FEB-1999 (Rel. 38, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Chalcone synthase E (EC 2.3.1.74) (Naringenin-chalcone synthase E)			
DE	(CHS-E).			
GN	CHSE.			
OS	Ipomoea purpurea (Common morning glory).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
OC	Assteridae; lamids; Solanales; Convolvulaceae; Ipomoea.			
OX	NCBI_TaxId=4121;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV, FR-35; TISSUE=flower buds;			
RX	MEDLINE=97393456; PubMed=9249950;			
RA	Fukuda-Tanaka S., Hoshino A., Hataomi Y., Habu Y., Hasebe M.,			
RT	Iida S.;			
RL	"Identification of new chalcone synthase genes for flower pigmentation			
RT	in the Japanese and common morning glories.";			
CC	Plant Cell Physiol. 38:754-756(1997).";			
CC	-1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-			
CC	tetrahydrochalcone (also termed naringenin-chalcone or chalcone)			
CC	which can under specific conditions spontaneously isomerize into			
CC	naringenin.			
CC	-1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +			
CC	naringenin-chalcone + 3 CO(2).			
CC	-1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF			
CC	FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY			
CC	OF WHICH ARE BRIGHTLY COLORED.			
CC	-1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
EMBL	AB001827; BAA21789.1; -			
PIR	T07799; T07799.			
DR	InterPro: IPR001099; N-C_ynthase.			
DR	Pfam: PF00195; Chal_sil_ynthc; 1.			
DR	Pfam: PF02797; Chal_sil_ynthc; 1.			
DR	ProDom: PD000453; N-C_ynthase; 1.			
DR	PROSITE: PS00441; CHALCONE_SYNTH; 1.			
KW	Flavonoid biosynthesis; Transferase; Acyltransferase;			
FT	MultiSub family.			
ACT_SITE	164			
FT	BY SIMILARITY			
SEQUENCE	389 AA; 42677 MW; 0CC8102CCA6A56A CRC64;			

Query Match	38.5%	Score 52	DB 1	Length 369
Best Local Similarity	64.3%	Pred. No. 2.6		
Matches	9	Conservative	3	Mismatches 2, Indels 0, Gaps 0
QY	5	OKAMKEWIOPKTKV	18	
			:	
DB	111	OKAIKEWQPKSKI	124	

```

RESULT 14
CHSE_PHANI STANDARD: PRT, 369 AA.
ID CHSE_PHANI 022046:
AC 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase B (EC 2.3.1.74) (Naringenin-chalcone synthase E)
DS (CHS-B).
GN CHSE.
OS Phorbite nil (Violet) (Japanese morning glory).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Convolvulaceae; Ipomoea.
CX NCBI_TaxID=35883;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. KK/ZSK-2; TISSUE=flower buds;
RX MEDLINE=97393496; PubMed=9249990;
RA Fukuda-Tanaka S., Hoshino A., Hataomi Y., Habu Y., Hasebe M.,
RA Iida S.;
RT "Identification of new chalcone synthase genes for flower pigmentation
RL in the Japanese and common morning glories.";
RL Plant Cell Physiol. 38:754-758(1997).
CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
CC tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS. A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL; AB001819; BAA21788.1; -
DR InterPro; IPR001099; N-C-synthase.
DR Pfam; PF00195; Chal_sbl_synth.
DR Pfam; PF02797; Chal_sbl_synth; 1.
DR ProDom; PD000453; N-C-synthase; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase;
KW Multigene family.
FT ACT_SITE 164 164 BY SIMILARITY.
SQ SEQUENCE 389 AA; 42685 MW; 0FC0E014B9CC0312 CRC64;
Qy Query Match 38.5%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 2.6;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Cy 5 OKAMKRWIOPTKRV 18
Db 111 OKAIKRWGQPSKI 124
RESULT 15
CHSJ_PETHY
ID CHSJ_PETHY STANDARD: PRT, 369 AA.
AC P22928;
DT 01-APR-1991 (Rel. 19, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase J (EC 2.3.1.74) (Naringenin-chalcone synthase J).
GN CHSJ.

```

```

OS  Petunia hybrida (Petunia).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC  Asteridae; Lamiales; Solanales; Solanaceae; Petunia.
OX  NCBI_TaxID=4102;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=cv. Violet 30; TISSUE=leaf;
RX  MEDLINE=90034197; PubMed=2806915;
RA  Koele R.E., Spelt C.E., van den Elzen P.J.M., Mol J.N.M.;
RT  "Cloning and molecular characterization of the chalcone synthase
RT  multigene family of Petunia hybrida.";
RL  Gene 81:245-257(1989).
RN  [2]
RP  SEQUENCE OF 71-389 FROM N.A.
RC  STRAIN=cv. White 137; TISSUE=anther;
RA  van Tunen A.J.;
RL  Submitted (MAR-1989) to the EMBL/GenBank/DBJ databases.
CC  -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
CC  tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
CC  which can under specific conditions spontaneously isomerize into
CC  naringenin.
CC  -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC  naringenin-chalcone + 3 CO(2).
CC  -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC  FLAVONOIDS. A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC  OF WHICH ARE BRIGHTLY COLORED.
CC  -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC
CC  -----
DR  EMBL; X14597; CAA32737.1; -.
DR  EMBL; X14599; CAA32739.1; -.
DR  PIR; D72821; SYPUCJ.
DR  PIR; S18136; S18136.
DR  InterPro; IPR001099; N-C_synthase.
DR  Pfam; PF00195; Chal_stil_synth; 1.
DR  pfam; PF02797; Chal_stil_synth; 1.
DR  ProDom; PD000453; N-C_synthase; 1.
DR  PROSITE; PS00441; CHALCONE_SYNTH; 1.
KW  Flavonoid biosynthesis; transferase; Acyltransferase;
KW  Multigene family.
FT  ACT SITE 164 164
FT  CONFLICT 75 75 E -> V (in Ref. 2).
SQ  SEQUENCE 389 AA; 42558 MW; F2B3CDD82E6FDE7D CRC64;

```

Query Match

Best Local Similarity 38.5%; Score 52; DB 1; Length 389;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMPMIOPKTV 18
 |||:|:|:|:|:
 Db 111 OKAIKMGOPKSKI 124

Search completed: July 30, 2003, 16:24:28
 Job time : 7.85714 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:55 ; Search time 15.2857 Seconds
(without alignments)
150.994 Million cell updates/sec

Title: US-09-787-070-3

Perfect score: 135

Sequence: 1 VYOHKAKMPWIPKTKVLPYVRY 24

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	135	100.0	222	1 KABOS2	alpha-s2-casein pr
2	111	82.2	223	1 KASHS2	alpha-s2-casein pr
3	111	82.2	223	2 JN0547	alpha-s2-casein pr
4	111	82.2	223	2 JN0547	alpha-s2-casein C
5	70.5	52.2	235	2 A48383	alpha-s2-casein -
6	67	49.6	182	2 S39776	alpha-s2-casein fo
7	53	39.3	935	2 T33389	probable DNA damag
8	52	38.5	349	2 S18136	naringenin-chalcon
9	52	38.5	340	2 T10957	naringenin-chalcon
10	52	38.5	389	1 SYRJCJ	naringenin-chalcon
11	52	38.5	389	1 SYRJCJ	naringenin-chalcon
12	52	38.5	389	1 SYRJCJ	naringenin-chalcon
13	52	38.5	389	1 SYRJCJ	naringenin-chalcon
14	52	38.5	389	1 SYRJCJ	naringenin-chalcon
15	52	38.5	389	1 SYRJCJ	naringenin-chalcon
16	52	38.5	389	1 SYRJCJ	naringenin-chalcon
17	50	37.0	1877	2 T10550	alpha-s2-casein fo
18	49.5	36.7	325	2 C84000	hypothetical prote
19	49.5	36.7	325	2 C84000	hypothetical prote
20	49.5	36.7	325	2 C84000	probable zinc meta
21	49.5	36.7	325	2 C84000	probable zinc meta
22	49.5	36.7	325	2 C84000	probable zinc meta
23	49.5	36.7	325	2 C84000	probable zinc meta
24	49.5	36.7	325	2 C84000	probable zinc meta
25	49.5	36.7	325	2 C84000	probable zinc meta
26	49.5	36.7	325	2 C84000	probable zinc meta
27	49.5	36.7	325	2 C84000	probable zinc meta
28	49.5	36.7	325	2 C84000	probable zinc meta
29	49.5	36.7	325	2 C84000	probable zinc meta

30	48	35.6	418	2 F72370	phoH-related prote
31	47	34.8	53	2 S41957	naringenin-chalcon
32	47	34.8	128	2 S33611	naringenin-chalcon
33	47	34.8	158	2 B86169	hypothetical prote
34	47	34.8	285	2 S44369	naringenin-chalcon
35	47	34.8	331	2 JQ1071	naringenin-chalcon
36	47	34.8	370	2 S44368	naringenin-chalcon
37	47	34.8	383	2 S31655	naringenin-chalcon
38	47	34.8	388	1 SYSCVN	naringenin-chalcon
39	47	34.8	388	1 SYSCVN	naringenin-chalcon
40	47	34.8	388	1 SYSCVN	naringenin-chalcon
41	47	34.8	388	1 SYSCVN	naringenin-chalcon
42	47	34.8	388	2 JQ2249	naringenin-chalcon
43	47	34.8	388	2 JQ2259	naringenin-chalcon
44	47	34.8	388	2 S60472	naringenin-chalcon
45	47	34.8	388	2 JCS516	naringenin-chalcon

ALIGNMENTS

RESULT 1
KABOS2
alpha-s2-casein precursor - bovine
C.Species: Bos primigenius taurus (cattle)
C.Date: 01-Sep-1981 #sequence revision 03-Feb-1994 #text_change 22-Jun-1999
C.Accession: JQ2008; A29087; A91438; S66628; A03107
R.Gronen, M.A.M.; Dijkhof, R.J.M.; Verstege, A.V.M.; van der Poel, J.J.
Gene 123, 187-193, 1993
A.Title: The complete sequence of the gene encoding bovine alpha-s2-casein.
A.Reference number: JQ2008; MUID:93154583; PMID:8428658
A.Accession: JQ2008
A.Status: translation not shown
A.Molecule type: DNA
A.Residues: 1-222 <GRO>
A.Cross-references: GB:M94327
R.Stewart, A.F.; Bomsing, J.; Beattie, C.W.; Shah, F.; Willis, I.M.; Mackinlay, A.G.
Mol. Biol. Evol. 4, 231-241, 1987
A.Title: Complete nucleotide sequences of bovine alpha-s2- and beta-casein cDNAs: compar
A.Reference number: A93062; MUID:88189899; PMID:2833669
A.Accession: A29087
A.Status: translation not shown
A.Molecule type: mRNA
A.Residues: 1-222 <STB>
A.Cross-references: GB:M16644; MUID:9162928; PID:AAA30479.1; PID:9162929
R.Brignon, G.; Ribadeau Dumas, B.; Mercier, J.C.; Pellesier, J.P.; Dae, B.C.
FEBS Lett. 76, 274-279, 1977
A.Title: Complete amino acid sequence of bovine alpha-s2-casein.
A.Reference number: A91438; MUID:77185633; PMID:862506
A.Accession: A91438
A.Contents: A allele
A.Molecule type: protein
A.Residues: 16-101, 'EE', 104-222

A.Note: four fractions, previously designated s2, s3, s4, and s6, appear to have the sam
these
R.Grosclaude, F.; Joudrier, P.; Mahe, M.F.
J. Dairy Res. 46, 211-213, 1979
A.Title: A genetic and biochemical analysis of a polymorphism of bovine alpha-s2-casein
A.Reference number: A92771; MUID:79339837; PMID:469044
A.Contents: annotation; D allele
A.Note: the sequence of the D allele has a deletion of nine residues, which may be 49-58
R.Zucht, H.D.; Rada, M.; Adermann, K.; Maegert, H.J.; Forsmann, W.G.
FEBS Lett. 372, 181-188, 1995
A.Title: Casocidin-1: a casein-alpha(s2) derived peptide exhibits antibacterial activity
A.Reference number: S66626; MUID:96000204; PMID:755666
A.Accession: S66626
A.Molecule type: protein
A.Residues: 165-203 <ZDC>
C.Comment: The sequence of the A allele is shown.
C.Genetics:
A.Gene: alpha2ca
A.Map position: 6
A.Introns: 17/3; 26/3; 33/3; 47/3; 56/3; 65/3; 74/3; 82/3; 97/3; 138/3; 147/3; 156/3; 16/

C/Superfamily: alpha-s2-casein
 C/Keywords: mammary gland; milk; phosphoprotein
 F/1-15/Domain: signal sequence #status predicted <SIG>
 F/16-222/Product: alpha-s2-casein #status experimental <MAT>
 F/23,24,25,31,71,72,73,76,144,146,158/Binding site: phosphate (Ser) (covalent) #status F

Query Match 100.0%; Score 135; DB 1; Length 222;
 Best Local Similarity 100.0%; Pred. No. 1,1e-12;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIQPKTKVLPYRY 24
 |||||
 198 VYQHOKAMKPMIQPKTKVLPYRY 221

RESULT 2

KASHS2
 alpha-s2-casein precursor - sheep
 C/Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C/Date: 31-Mar-1988 #sequence_revansion 31-Mar-1988 #text_change 22-Jun-1999
 C/Accession: A25070; S17856
 R/Boulton, M.; Petrisant, G.
 Biochimie 67, 1043-1051, 1985
 A/Title: Complete sequence of ovine alpha-s2-casein messenger RNA.
 A/Reference number: A25070; MUID:86104467; PMID:3002499
 A/Accession: A25070

A/Molecule type: mRNA
 A/Residues: 1-223 <BO1>
 A/Cross-references: GB:X03228; NID:g1238; PIDN:CAA26983.1; PID:g732894
 A/Note: 64-Aan was also found
 R/Boulton, M.; Hue, D.; Boulton, C.; Mercier, J.C.; Gaye, P.
 Eur. J. Biochem. 201, 633-641, 1991
 A/Title: Multiple RNA species code for two non-allelic forms of ovine alpha-s2-casein.
 A/Reference number: S17856; MUID:92037619; PMID:1935859

A/Accession: S17856
 A/Status: Preliminary
 A/Molecule type: mRNA
 A/Residues: 1-93, 'T', '95-223 <BO2>
 C/Superfamily: alpha-s2-casein
 C/Keywords: mammary gland; milk; phosphoprotein
 F/1-15/Domain: signal sequence #status predicted <SIG>
 F/16-223/Product: alpha-s2-casein #status predicted <KA>
 F/23,24,25,32,55,72,73,74,77,145,147,159/Binding site: phosphate (Thr) (covalent) #status F/23,24,25,32,55,72,73,74,77,145,147,159/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 82.2%; Score 111; DB 1; Length 223;
 Best Local Similarity 83.3%; Pred. No. 4,1e-09;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIQPKTKVLPYRY 24
 |||||
 199 VYQHOKAMKPMIQPKTKVLPYRY 222

RESULT 3

JN0547
 alpha-s2-casein precursor - goat
 C/Species: Capra aegagrus hircus (domestic goat)
 C/Date: 31-Dec-1993 #sequence_revansion 15-Oct-1999 #text_change 15-Oct-1999
 C/Accession: S33881; S33880; JN0547; S20620
 R/Boulton, C.; Brignon, G.; Mahe, M.F.; Printz, C.
 Protein Seq. Data Anal. 5, 213-218, 1993
 A/Title: Characterization of goat allelic alpha-s2-caseins A and B: further evidence of

A/Reference number: S33880
 A/Accession: S33881
 A/Status: preliminary
 A/Molecule type: protein
 A/Residues: 1-223 <BO1>
 A/Note: allele B
 A/Accession: S33880
 A/Molecule type: protein
 A/Residues: 16-78, 'E', '80-223 <BO2>
 A/Note: sequence deduced from compositional analysis of peptides

R/Boulton, C.
 Gene 125, 235-236, 1993
 A/Title: Sequence of the goat alpha-s2-casein-encoding cDNA.
 A/Reference number: JN0547; MUID:93216130; PMID:8462880
 A/Accession: JN0547

A/Molecule type: mRNA
 A/Residues: 1-78, 'E', '80-223 <BO3>
 A/Cross-references: EMBL:X65160; NID:g955; PIDN:CAA46278.1; PID:g956
 A/Note: allele A
 C/Superfamily: alpha-s2-casein
 C/Keywords: mammary gland; milk; phosphoprotein
 F/1-15/Domain: signal sequence #status predicted <SIG>
 F/16-223/Product: alpha-s2-casein #status experimental <MAT>
 F/23,24,25,55,72,73,74,77,145,147,159/Binding site: phosphate (Ser) (covalent) #status F/23,24,25,55,72,73,74,77,145,147,159/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 82.2%; Score 111; DB 2; Length 223;
 Best Local Similarity 83.3%; Pred. No. 4,1e-09;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIQPKTKVLPYRY 24
 |||||
 199 VYQHOKAMKPMIQPKTKVLPYRY 222

RESULT 4

I46995
 alpha s2-casein C - goat
 C/Species: Capra aegagrus hircus (domestic goat)
 C/Date: 21-Feb-1997 #sequence_revansion 21-Feb-1997 #text_change 13-Aug-1999
 C/Accession: I46995
 R/Boulton, C.; Brignon, G.; Mahe, M.F.; Printz, C.
 Anim. Genet. 25, 173-177, 1994
 A/Title: Biochemical and genetic analysis of variant C of caprine alpha s2-casein (Capr

A/Reference number: I46995; MUID:95030556; PMID:7943951
 A/Accession: I46995
 A/Status: Preliminary; translated from GB/EMBL/DDBU
 A/Molecule type: DNA
 A/Residues: 1-223 <BO1>
 A/Cross-references: GB:S74171; NID:g707033; PIDN:AAB32166.1; PID:g707034
 C/Superfamily: alpha-s2-casein

Query Match 82.2%; Score 111; DB 2; Length 223;
 Best Local Similarity 83.3%; Pred. No. 4,1e-09;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIQPKTKVLPYRY 24
 |||||
 199 VYQHOKAMKPMIQPKTKVLPYRY 222

RESULT 5

A48383
 alpha s2-casein - pig
 C/Species: Sus scrofa domestica (domestic pig)
 C/Date: 19-Nov-1993 #sequence_revansion 18-Nov-1994 #text_change 03-May-1996
 C/Accession: A48383
 R/Alexander, L.J.; Das Gupta, N.A.; Beattie, C.W.
 Anim. Genet. 23, 365-367, 1992

A/Title: The sequence of porcine alpha s2-casein cDNA.
 A/Reference number: A48383; MUID:92367960; PMID:1503276
 A/Accession: A48383
 A/Status: preliminary
 A/Molecule type: nucleic acid
 A/Residues: 1-235 <ALB>
 A/Experimental source: mammary gland
 A/Note: sequence inconsistent with the nucleotide translation
 A/Note: sequence extracted from NCBI backbone (NCBIN:110884; NCBI:P:110885)
 C/Superfamily: alpha-s2-casein

Query Match 52.2%; Score 70.5; DB 2; Length 235;
 Best Local Similarity 55.6%; Pred. No. 0.005;
 Matches 15; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

A:Gene: chs1
A:Map position: V
A:Note: chs1 is expressed in various floral tissues and UV illuminated seedlings
C:Superfamily: chalcone synthase
C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 38.5%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 5;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKMGQPKSKI 124

RESULT 11

SYBJCN
naringenin-chalcone synthase (EC 2.3.1.74) R - garden petunia

N:Alternate names: chalcone synthase
C:Species: Petunia x hybrida (garden petunia)

C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000
C:Accession: A23643

R:Koebe, R.E.; Spelt, C.E.; Reif, H.J.; van den Elzen, P.J.M.; Veltkamp, E.; Mol, J.N.M.
Nucleic Acids Res. 14, 5229-5239, 1986

A:Title: Floral tissue of Petunia hybrida (V30) expresses only one member of the chalcone
A:Reference number: A23643; MUID:66286540; PMID:3016642

A:Accession: A23643
A:Molecule type: mRNA

A:Residues: 1-389 <KOE>
A:Cross-references: GB:X04080; NID:G20541; PIDN:CAA27711.1; PID:G20542

A:Experimental source: strain Violet 30, flowers
C:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavonoid

C:Genetics:
A:Gene: chs1
A:Note: expressed in floral tissue

C:Superfamily: chalcone synthase
C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 38.5%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 5;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKMGQPKSKI 124

RESULT 12

SYBJCN

naringenin-chalcone synthase (EC 2.3.1.74) A - garden petunia

N:Alternate names: chalcone synthase
C:Species: Petunia x hybrida (garden petunia)

C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000
C:Accession: J50308

R:Koebe, R.E.; Spelt, C.E.; van den Elzen, P.J.M.; Mol, J.N.M.
Gene 81, 245-257, 1989

A:Title: Cloning and molecular characterization of the chalcone synthase multigene family
A:Reference number: J50308; MUID:90034197; PMID:2806915

A:Accession: J50308
A:Molecule type: DNA

A:Residues: 1-389 <KOE>
A:Cross-references: GB:X14591; NID:G20524; PIDN:CAA37731.1; PID:G20525

A:Experimental source: strain Violet 30, leaf
C:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavonoid

C:Genetics:
A:Gene: chs1

A:Map position: V
A:Note: chs1 is the major expressed member of the gene family in various floral tissues A

C:Superfamily: chalcone synthase
C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 38.5%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 5;

Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKMGQPKSKI 124

RESULT 13

J50316

naringenin-chalcone synthase (EC 2.3.1.74) 2 - potato

N:Alternate names: chalcone synthase; CHS
C:Species: Solanum tuberosum (potato)

C>Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 05-May-2000
C:Accession: J50316; PC4239

R:Jeon, J.H.; Kim, H.S.; Choi, K.H.; Jeong, Y.H.; Jeong, H.; Byun, S.M.
Biosci. Biotechnol. Biochem. 60, 1907-1910, 1996

A:Title: Cloning and characterization of one member of the chalcone synthase gene family
A:Reference number: J50316; MUID:97141614; PMID:8987872

A:Accession: J50316
A:Status: preliminary

A:Molecule type: mRNA
A:Residues: 1-389 <JEO1>

A:Cross-references: GB:U47738; NID:G1470059; PIDN:AA05239.1; PID:G1470060
A:Accession: PC4239

A:Status: preliminary
A:Molecule type: protein

A:Residues: 158-165;567-573 <JEO2>
C:Comment: This enzyme is important in the biosynthesis of all classes of flavonoids in

C:Superfamily: chalcone synthase
C:Keywords: acyltransferase; coenzyme A

Query Match 38.5%; Score 52; DB 2; Length 389;
Best Local Similarity 64.3%; Pred. No. 5;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKMGQPKSKI 124

RESULT 14

T07799
naringenin-chalcone synthase (EC 2.3.1.74) - common morning-glory

N:Alternate names: chalcone synthase
C:Species: Ipomoea purpurea (common morning-glory)

C>Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
C:Accession: T07799

R:Fukuda-Tanaka, S.; Hoshino, A.; Hisatomi, Y.; Habu, Y.; Hasebe, M.; Iida, S.
Plant Cell Physiol. 38, 754-759, 1997

A:Title: Identification of new chalcone synthase genes for flower pigmentation in the J

A:Reference number: Z16140; MUID:97383496; PMID:9249990
A:Accession: T07799

A:Status: preliminary; translated from GB/EMBL/DDBU
A:Molecule type: mRNA

A:Residues: 1-389 <FKU>
A:Cross-references: EMBL:AB001827; NID:G2329836; PIDN:BAA21789.1; PID:G2329837

C:Genetics:
A:Gene: CHD-E

C:Superfamily: chalcone synthase
C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 38.5%; Score 52; DB 2; Length 389;
Best Local Similarity 64.3%; Pred. No. 5;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKMGQPKSKI 124

RESULT 15

SYBKCD
naringenin-chalcone synthase (EC 2.3.1.74) - garden snapdragon

N:Alternate names: chalcone synthase
 C:Species: Antirrhinum majus (garden snapdragon)
 C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000
 C:Accession: S07312; A33217
 R:Sommer, H.; Saedler, H.
 M:J. Gen. Genet. 202, 429-434, 1986
 A:Title: Structure of the chalcone synthase gene of Antirrhinum majus.
 A:Reference number: S07312
 A:Accession: S07312
 A:Molecule type: DNA
 A:Residues: 1-390 <SOM>
 A:Cross-references: EMBL:X037710; NID:G16015; PID:CAA27338.1; PID:G16016
 A:Accession: A33217
 A:Status: translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-390 <SOM2>
 C:Genetics:
 A:Gene: chs
 A:Introns: 60/1; 162/3
 C:Superfamily: chalcone synthase
 C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 38.5%; Score 52; DB 1; Length 390;
 Best Local Similarity 64.3%; Pred. No. 5;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 OKMKPMTOPRTKV 18
 |||:|||:
 Db 111 OKAIKMGQPSKSI 124

Search completed: July 30, 2003, 16:31:15
 Job time : 15.2857 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: July 30, 2003, 16:09:50 ; Search time 47.2857 Seconds
(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-3
Perfect score: 135
Sequence: 1 VYOHQKAKKFWIQKTVIPYRV 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	135	100.0	24	15 AAR60481	Peptide used in tr
2	135	100.0	24	21 AAY88338	Internal peptide f
3	135	100.0	25	13 AAR20055	Casain peptide. S
4	135	100.0	25	21 AAY88340	C-terminal peptide
5	135	100.0	31	18 AAW32219	Alpha-S2 casein pr
6	135	100.0	31	23 AAE17466	Bovine alpha-S2 ca
7	135	100.0	222	18 AAW32220	Bovine alpha-S2 ca
8	135	100.0	222	23 AAE17468	Bovine alpha-S2 ca
9	124.5	92.2	26	20 AAW93885	Bifidobacterium bi

10	119	88.1	21	21 AAY88338	Internal peptide f
11	111	82.2	23	23 AAE17469	Alpha-S2 casein pr
12	111	82.2	23	23 AAE17470	Goat alpha-S2 case
13	111	82.2	23	23 AAE17471	Capra hircus alpha
14	111	82.2	23	23 AAE17473	Sheep alpha-S2 cas
15	111	82.2	23	23 AAE17475	Alpha-S2 casein pr
16	101	74.8	19	18 AAW32218	Alpha-S2 casein pr
17	101	74.8	19	23 AAE17465	Bovine alpha-S2 ca
18	92	68.1	16	21 AAY88337	Internal peptide f
19	70.5	52.2	24	23 AAE17472	Pig alpha-S2 casei
20	67	49.6	178	23 AAE17474	Rabbit alpha-S2 ca
21	65	48.1	13	23 AAE17464	Bovine alpha-S2 ca
22	60	44.4	12	23 AAE17463	Bovine alpha-S2 ca
23	53	39.3	9	17 AAW04282	Milk derived anti-
24	53	39.3	11	18 AAW32217	Alpha-S2 casein pr
25	53	39.3	11	23 AAE17462	Bovine alpha-S2 ca
26	52	38.5	360	22 AAB60169	Petunia chalcone s
27	52	38.5	389	17 AAB95125	Chalcone synthase.
28	52	38.5	388	22 AAB74019	Cyclamen chalcone
29	52	38.5	389	24 AAB73901	Petunia chalcone s
30	51	37.8	394	24 AAB60374	Humulus lupulus ch
31	51	37.8	399	24 AAB60375	Humulus lupulus ch
32	50	37.0	291	22 AAB69092	Mouse GM protein s
33	49	36.3	8	17 AAW04278	Milk derived anti-
34	49	36.3	133	17 AAW04828	Vascular endotheli
35	49	36.3	133	20 AAW80492	Vascular endotheli
36	49	36.3	133	23 AAB83406	Marine vascular en
37	49	36.3	160	20 AAW86208	Human VEGF-B trunc
38	49	36.3	167	20 AAW86234	Human VEGF-B full
39	49	36.3	188	17 AAW86234	Heart vascular end
40	49	36.3	188	17 AAW00864	Human VRF167. Ma
41	49	36.3	188	20 AAW86201	Marine vascular en
42	49	36.3	188	20 AAW80490	Marine vascular en
43	49	36.3	188	23 AAB83404	Marine vascular en
44	49	36.3	195	17 AAW04827	Heart vascular end
45	49	36.3	195	20 AAW80491	Marine vascular en

ALIGNMENTS

RESULT 1
ID AAR60481 standard; peptide; 24 AA.
AAR60481;
19-APR-1995 (first entry)
Peptide used in tranquilliser.
Peptide; tranquilliser; food; foodstuff; mental stabilisation.
Synthetic.
JP06211689-A.
02-AUG-1994.
19-JAN-1993; 93JP-0024811.
19-JAN-1993; 93JP-0024811.
(KANE) KANEBO LTD.
WPI; 1994-283276/35.
Synthetic peptide(s) used in tranquilliser - also used in the production of a food for mental stabilisation.
Claim 3; Page 1; 5pp; Japanese.
The peptide may be used as a component of a tranquilliser which may

CC be administered orally and is low in toxicity. The peptide may also
CC be used as a component of a foodstuff which may be used for mental
CC stabilisation. See also AAR60480.

XX Sequence 24 AA;

Query Match 100.0%; Score 135; DB 15; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.2e-13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPTKVIPIYVRY 24
DB 1 VYOHOKAMKPMIOPTKVIPIYVRY 24

RESULT 2
ID AAY88339 standard; peptide; 24 AA.

XX AAY88339;

DT 14-JUL-2000 (first entry)

DE Internal peptide fragment of bovine alpha-S2 caesin #3.

KM Alpha-S2 caesin; peptide production; biological fluid; milk; whey; blood;
XX antibacterial peptide; lactoferrin; antiviral; antitumour activity.

OS Bos sp.

PN WO200015655-A1.

PD 23-MAR-2000.

PF 15-SEP-1999; 99WO-EP07002.

PR 15-SEP-1998; 98EP-0203107.

PR 08-JUN-1999; 99EP-0201815.

PA (NIZO-) NIZO FOOD RES.

PI Visser S, Recto I;

DR WPI; 2000-271377/23.

PT Novel process for producing peptides with e.g. antimicrobial activity
PT from biological fluids such as milk, whey or blood comprises contacting
PT fluid with chromatographic medium to adsorb peptide domain of interest

PS Claim 14; Page 22; 41pp; English.

CC This sequence represents an internal fragment of bovine alpha-S2 caesin
CC protein. The peptide is an example of a peptide with antibacterial
CC activity that can be produced by the process of the invention. The
CC invention relates to a process for producing peptides from biological
CC fluids. The process comprises chromatography of the biological fluid, in
CC situ hydrolysis of selectively bound peptides, washing to remove unbound
CC peptide, and elution of the peptides of interest. The process is used for
CC producing peptides from biological fluids, such as milk, whey or blood.
CC For example, the process can be used to produce antibacterial peptides
CC derived from lactoferrin, using cheese whey as a starting material. The
CC peptides obtained have preferably antimicrobial and/or antiviral and/or
CC antitumour activity. The process of the invention is relatively simple
CC and generally economically and technically more attractive than those
CC methods previously used. The method provides high yield peptide with a
CC selected activity of interest without the need for intermediate
CC purification of the precursor protein.

XX Sequence 24 AA;

Query Match 100.0%; Score 135; DB 21; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.2e-13;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPTKVIPIYVRY 24
DB 1 VYOHOKAMKPMIOPTKVIPIYVRY 24

RESULT 3
ID AAR20055 standard; peptide; 25 AA.

XX AAR20055;

DT 25-MAR-2003 (updated)

DT 26-MAR-1992 (first entry)

DE Casein peptide.

KM Casein; platelet; aggregation; thrombosis; collagen.

OS Synthetic.

PN JP03255095-A.

PD 13-NOV-1991.

PF 02-MAR-1990; 90JP-0052554.

PR 02-MAR-1990; 90JP-0052554.

PA (KANE) KANEBO LTD.

DR WPI; 1992-002669/01.

PT Casein peptide(s) for treating thrombosis - as inhibitors of
PT platelet aggregation caused by adenosine-5'-diphosphate and
PT collagen and as biochemical reagents

PS Claim 1; Page 1; 8pp; Japanese.

CC Leu25 may be omitted. The peptide and its salts have inhibiting
CC activity against platelet aggregation caused by adenosine-5'-
CC diphosphate and by collagen. They are useful for prevention and
CC treatment of thrombosis and are also useful as a biochemical reagent.
CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 25 AA;

Query Match 100.0%; Score 135; DB 13; Length 25;

Best Local Similarity 100.0%; Pred. No. 6.5e-13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPTKVIPIYVRY 24
DB 1 VYOHOKAMKPMIOPTKVIPIYVRY 24

RESULT 4
ID AAY88340 standard; peptide; 25 AA.

XX AAY88340;

DT 14-JUL-2000 (first entry)

DE C-terminal peptide fragment of bovine alpha-S2 caesin.

KM Alpha-S2 caesin; peptide production; biological fluid; milk; whey; blood;

KM antibacterial peptide; lactoferrin; antiviral; antitumour activity.

OS Bos sp.

PN WO200015655-A1.

XX	23-MAR-2000.	
XX	15-SEP-1999;	99WO-EP07002.
XX	15-SEP-1998;	98EP-0203107.
XX	08-JUN-1999;	99EP-0201815.
XX	(NIZO-) NIZO FOOD RES.	
XX	Vlaesser S, Recio I;	
XX	WPI, 2000-271377/23.	
XX	Novel process for producing peptides with e.g. antimicrobial activity	
XX	from biological fluids such as milk, whey or blood comprises contacting	
XX	fluid with chromatographic medium to adsorb peptide domain of interest	
XX	Claim 14; Page 22; 41pp; English.	
XX	This sequence represents a C-terminal fragment of bovine alpha-S2 caesin	
XX	protein. The peptide is an example of a peptide with antibacterial	
XX	activity that can be produced by the process of the invention. The	
XX	invention relates to a process for producing peptides from biological	
XX	fluids. The process comprises chromatography of the biological fluid, in	
XX	situ hydrolysis of selectively bound peptides, washing to remove unbound	
XX	peptide, and elution of the peptides of interest. The process is used for	
XX	producing peptides from biological fluids, such as milk, whey or blood.	
XX	For example, the process can be used to produce antibacterial peptides	
XX	derived from lactoferrin, using cheese whey as a starting material. The	
XX	peptides obtained have preferably antimicrobial and/or antiviral and/or	
XX	antitumour activity. The process of the invention is relatively simple	
XX	and generally economically and technically more attractive than those	
XX	methods previously used. The method provides high yield peptides with a	
XX	selected activity of interest without the need for intermediate	
XX	purification of the precursor protein.	
XX	Sequence 25 AA;	
XX	Query Match 100.0%; Score 135; DB 21; Length 25;	
XX	Best Local Similarity 100.0%; Pred. No. 6,5e-13;	
XX	Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
XX	1 VYOHQAMKPMWIOPTKTVIPVRY 24	
XX	1 VYOHQAMKPMWIOPTKTVIPVRY 24	
XX	Db	
XX	RESULT 5	
XX	AAW32219	
XX	AAW32219 standard; peptide; 31 AA.	
XX	AAW32219;	
XX	03-FEB-1998 (first entry)	
XX	Alpha-S2 casein precursor C-terminal peptide fragment #5.	
XX	Alpha-S2 casein precursor; growth promoting; mitogenic assay;	
XX	platelet-derived growth factor; insulin-like derived growth factor;	
XX	Synthetic.	
XX	Bos taurus.	
XX	MO9716460-A1.	
XX	09-MAY-1997.	
XX	31-OCT-1996;	96WO-GB02658.
XX	31-OCT-1995;	95GB-0022302.

PA (UYLI-) UNTY LIVERPOOL.
XX
PI Liu Q, Smith JA, Wilkinson MC;
XX
XX WPI; 1997-272048/24.
DR
XX
XX Manufacture of medicament or foodstuff for promoting growth - using
PT peptide(s) with a sequence identical to the C-terminal end of an
PT alpha-S2 casein precursor
XX
PS Claim 12; Page 21; 33p; English.
XX
XX The present sequence, which is substantially identical to the C-terminal
CC end of an alpha-S2 casein precursor, was found after storage in PBS to
CC exhibit growth promoting activity for rat mammary fibroblast cell line
CC (Rama 27), which is not significantly stimulated by platelet-derived
CC growth factor or insulin-like growth factor. The activity of the peptide
CC increased when maintained at alkaline pH. By way of contrast, alpha-S2
CC casein was inactive in a mitogenic assay. This peptide may be used in the
CC manufacture of a medicament or foodstuff for promoting growth in
CC humans or animals.
XX
SQ Sequence 31 AA;
XX
Query Match 100.0%; Score 135; DB 18; Length 31;
Best Local Similarity 100.0%; Pred. No. 8.3e-13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 1 VYOHQKAMKFWIQPKTKVIPYRY 24
|||
7 VYOHQKAMKFWIQPKTKVIPYRY 30
DB
RESULT 6
AAE17466
ID AAE17466 strand; peptide; 31 AA.
AC
XX AAE17466;
AC
XX
DT 22-APR-2002 (first entry)
XX
XX Bovine alpha-S2 casein precursor protein C-terminal fragment #7.
DE
XX
XX Bovine; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
KM chewing gum; cosmetic; wrinkling; periodontal disease.
XX
OS Bos sp.
XX
XX WO200202133-A2.
XX
XX 10-JAN-2002.
XX
PD 13-JUN-2001; 2001WO-GB02601.
XX
XX 30-JUN-2000; 2000GB-0016189.
XX
XX (PEPS-) PEPSYN LTD.
XX
XX Smith JA;
XX
XX WPI; 2002-154690/20.
XX
XX
XX Use of peptide or its derivative containing an amino acid sequence in
PT alpha-S2 casein precursor in the manufacture of a medicament for
PT alleviating or preventing periodontal disease and an effect of aging in
PT skin -
XX
PS Claim 9; Page 22; 27p; English.

CC thus the synthesis and secretion of collagen. The peptides also stimulate
CC the growth of keratinocytes, which aid in the formation and regeneration
CC of skin surface. The peptide is useful in the manufacture of a medicament
CC in the form of a toothpaste or a chewing gum, for alleviating or
CC preventing periodontal disease and a medicament in the form of a cosmetic
CC composition for alleviating or preventing an effect of aging,
CC particularly wrinkling of the skin. The present sequence is a peptide
CC derived from the C-terminal of bovine alpha-S2 casein precursor protein.

XX Sequence 31 AA;

Query Match 100.0%; Score 135; DB 23; Length 31;
Best Local Similarity 100.0%; Pred. No. 8.3e-13;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VYQHOKAMKPMIOPKTVIPYRY 24
DB 7 VYQHOKAMKPMIOPKTVIPYRY 30

RESULT 7

AAM32220
ID AAM32220 standard; protein; 222 AA.

AC AAM32220;

DT 03-FEB-1998 (first entry)

DE Bovine alpha-S2 casein precursor.

KW Bovine alpha-S2 casein precursor; growth promoting; mitogenic assay;

KM platelet-derived growth factor; insulin-like derived growth factor;

XX Bos taurus.

XX WO9716460-A1.

XX 09-MAY-1997.

XX 31-OCT-1996; 96WO-GB02658.

XX 31-OCT-1995; 95GB-0022302.

PA (UPLI-) UNIV LIVERPOOL.

PI Liu Q, Smith JA, Wilkinson MC;

XX WPI; 1997-272048/24.

PT Manufacture of medicament or foodstuff for promoting growth - using
PT peptide(s) with a sequence identical to the C-terminal end of an
PT alpha-S2 casein precursor

XX Disclosure; Page 3; 33pp; English.

XX The present sequence represents bovine alpha-S2 casein precursor.

CC Peptides having an amino acid sequence which is substantially identical
CC to the C-terminal end of an alpha-S2 casein precursor, are used for

CC manufacture of a medicament or foodstuff for promoting growth in humans
CC or animals.

XX Sequence 222 AA;

Query Match 100.0%; Score 135; DB 18; Length 222;
Best Local Similarity 100.0%; Pred. No. 7.3e-12;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VYQHOKAMKPMIOPKTVIPYRY 24
DB 198 VYQHOKAMKPMIOPKTVIPYRY 221

RESULT 8

AAE17468
ID AAE17468 standard; protein; 222 AA.

XX AAE17468;

DT 22-APR-2002 (first entry)

DE Bovine alpha-S2 casein precursor protein.

KW Bovine; alpha-S2 casein; dermatological; anti-inflammatory; fibroblast;

KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;

KW chewing gum; cosmetic; wrinkling; periodontal disease.

XX Bos sp.

XX WO200202133-A2.

XX 10-JAN-2002.

XX 13-JUN-2001; 2001WO-GB02601.

XX 30-JUN-2000; 2000GB-0016189.

XX (PEPS-) PEPSYN LTD.

XX Smith JA;

XX WPI; 2002-154690/20.

XX Use of peptide or its derivative containing an amino acid sequence in
XX alpha-S2 casein precursor in the manufacture of a medicament for
XX alleviating or preventing periodontal disease and an effect of aging in
XX skin

XX Claim 8; Page 6; 27pp; English.

XX The invention relates to a composition comprising a peptide or its
XX derivative. The peptide contains an amino acid sequence from alpha-S2
XX casein precursor. The peptides stimulate the growth of fibroblasts, and
XX thus the synthesis and secretion of collagen. The peptides also stimulate
XX the growth of keratinocytes, which aid in the formation and regeneration
XX of skin surface. The peptide is useful in the manufacture of a medicament
XX in the form of a toothpaste or a chewing gum, for alleviating or cosmetic
XX preventing periodontal disease and a medicament in the form of a cosmetic
XX composition for alleviating or preventing an effect of aging,
XX particularly wrinkling of the skin. The present sequence is bovine
XX alpha-S2 casein precursor protein.

XX Sequence 222 AA;

Query Match 100.0%; Score 135; DB 23; Length 222;
Best Local Similarity 100.0%; Pred. No. 7.3e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VYQHOKAMKPMIOPKTVIPYRY 24
DB 198 VYQHOKAMKPMIOPKTVIPYRY 221

RESULT 9

AAW93885
ID AAW93885 standard; peptide; 26 AA.

XX AAW93885;

DT 25-JUN-1999 (first entry)

DE Bifidobacterium bifidus stimulating peptide 21.

XX Bifidogenic peptide; protease; treatment; microbe-related disease;

KW bacteria; fungi; yeast; protozoa; virus; mycoplasma; filaria; plasmodia;

KM infection; inflammation; microbial induced tumour; degenerative disorder;

KW diarrhoea; colic; oral microflora; intestinal microflora; caries;

KW vaginal microflora.
 OS Bifidobacterium bifidus.
 PN MO9914231-A2.
 XX
 PD 25-MAR-1999.
 XX
 PF 16-SEP-1998; 98MO-EP05899.
 XX
 PR 11-FEB-1998; 98DE-1005385.
 PR 16-SEP-1997; 97DE-1040604.
 XX
 PA (FORS/) FORSSMANN W.
 PI Foresmann W, Lapeke C, Zucht H;
 XX WPI; 1999-244022/20.
 DR
 PT Milk-derived peptides that stimulate Bifidobacterium bifidus
 XX
 PS Claim 2; Page 3; 25pp; German.
 CC This invention describes milk-derived bifidogenic peptides and their
 CC active derivatives or fragments, and combinations of them produced by
 CC chemical coupling. Such are produced from bovine or human milk by
 CC treatment for 2 hr with proteases, then centrifuging to remove fat and
 CC acidifying to pH 2 to precipitate proteins. The solution phase is then
 CC subjected to reverse-phase high-performance liquid chromatography (HPLC)
 CC and cation-exchange HPLC, the fractions adjusted to salt content below
 CC 25 mM (by dialysis or reverse-phase HPLC) and tested for activity by
 CC growing Bifidobacterium bifidus and Escherichia coli in presence of the
 CC fractions. Those fractions for which (Bw-B0) - (Bw-E0) is at least 0.15
 CC are selected where Bw = germ count after 16 hr culture of B. bifidus in
 CC 50% Elliker broth containing peptide at 0.2 mg/ml, B0 = germ count under
 CC similar conditions in a peptide-free control, Bw = germ count after 16
 CC hr culture of E. coli in 3 g/l tryptic broth containing peptide at
 CC 0.2 mg/ml, B0 = germ count under similar conditions in a peptide-free
 CC control. The peptides AA93865-W93888 are used to treat microbe-related
 CC diseases caused by bacteria, fungi, yeast, protozoa, viruses,
 CC mycoplasmas, filaria and plasmodia, e.g. infections, inflammation,
 CC microbially induced tumours or degenerative disorders, diarrhoea, colic,
 CC abnormalities in oral, intestinal or vaginal microflora, or caries.
 CC
 SQ Sequence 26 AA;
 XX
 Query Match 92.2%; Score 124.5; DB 20; Length 26;
 Best Local Similarity 96.0%; Pred. No. 2, 4e-11;
 Matches 24; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 VYOHQKAM-KPWIOPKTKVTPYRY 24
 DB 1 VYOHQKAMKPWIOPKTKVTPYRY 25
 XX
 RESULT 10
 AA98338
 ID AA98338 standard; peptide; 21 AA.
 AC AA98338;
 XX
 DT 14-JUN-2000 (first entry)
 XX
 DE Internal peptide fragment of bovine alpha-S2 casein #2.
 XX
 KW Alpha-S2 casein; peptide production; biological fluid; milk; whey; blood;
 KW antibacterial peptide; lactoferrin; antiviral; antitumour activity.
 XX
 OS Bos sp.
 XX
 PN WO200015655-A1.
 XX
 PD 23-MAR-2000.

XX
 PF 15-SEP-1999; 99MO-EP07002.
 XX
 PR 15-SEP-1998; 98EP-0203107.
 PR 08-JUN-1999; 99EP-0201815.
 XX
 PA (NIZO-) NIZO FOOD RES.
 XX
 PI Visser S, Recio I;
 XX WPI; 2000-271377/23.
 DR
 PT Novel process for producing peptides with e.g. antimicrobial activity
 PT from biological fluids such as milk, whey or blood comprises contacting
 PT fluid with chromatographic medium to adsorb peptide domain of interest
 PT
 PS Claim 14; Page 22; 41pp; English.
 XX
 CC This sequence represents an internal fragment of bovine alpha-S2 casein
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.
 CC
 SQ Sequence 21 AA;
 XX
 Query Match 88.1%; Score 119; DB 21; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1, 2e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYOHQKAMKPWIOPKTKVTPY 21
 DB 1 VYOHQKAMKPWIOPKTKVTPY 21
 XX
 RESULT 11
 AA917469
 ID AA917469 standard; Protein; 223 AA.
 AC AA917469;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Alpha-S2 casein precursor (alpha-S2-CN) protein #1.
 XX
 KW Alpha-S2 casein; alpha-S2-CN; dermatological; antiinflammatory; cosmetic;
 KW fibroblast; collagen; keratinocyte; skin regeneration; medicament; aging;
 KW toothpaste; chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Unidentified.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX

PI Smith JA;
 XX WPI; 2002-154690/20.
 DR
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 XX skin -
 XX
 PS Disclosure; Page 8; 27pp; English.
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptide stimulates the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is alpha-S2
 CC casein precursor (alpha-S2-CN) protein.
 CC
 SQ Sequence 223 AA;
 Query Match 82.2%; Score 111; DB 23; Length 223;
 Best Local Similarity 83.3%; Pred. No. 2.5e-08;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 VYOHQXAMKPMIOPTKVIPIYVRY 24
 DB 159 VDOHOKAMKPMIOPTKVIPIYVRY 222
 RESULT 12
 AAEL17470
 ID AAE17470 standard; Protein: 223 AA.
 AC AAE17470;
 DT 22-APR-2002 (first entry)
 XX
 DE Goat alpha-S2 casein E precursor protein.
 XX
 KM Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KM chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Capra hircus.
 XX
 PN WO200202133-A2.
 PD 10-JAN-2002.
 PF 13-JUN-2001; 2001WO-GB02601.
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 XX skin -
 XX
 PS Claim 8; Page 8; 27pp; English.
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2

CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat
 CC alpha-S2 casein E precursor protein.
 CC
 SQ Sequence 223 AA;
 Query Match 82.2%; Score 111; DB 23; Length 223;
 Best Local Similarity 83.3%; Pred. No. 2.5e-08;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 VYOHQXAMKPMIOPTKVIPIYVRY 24
 DB 159 VDOHOKAMKPMIOPTKVIPIYVRY 222
 RESULT 13
 AAEL17471
 ID AAE17471 standard; Protein: 223 AA.
 AC AAE17471;
 DT 22-APR-2002 (first entry)
 XX
 DE Capra hircus alpha-S2 casein C precursor protein.
 XX
 KM Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KM chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Capra hircus.
 XX
 PN WO200202133-A2.
 PD 10-JAN-2002.
 PF 13-JUN-2001; 2001WO-GB02601.
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 XX skin -
 XX
 PS Claim 8; Page 8; 27pp; English.
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat
 CC alpha-S2 casein C precursor protein.
 CC
 SQ Sequence 223 AA;
 Query Match 82.2%; Score 111; DB 23; Length 223;

Best Local Similarity 83.3%; Pred. No. 2.5e-08;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQXAMKPMWIOPTKVIPYRY 24
DB 199 VDOHQXAMKPMWIOPTKVTAIPYRY 222

RESULT 14

AAE17473
ID AAE17473 standard; Protein: 223 AA.

AC AAE17473;

DT 22-APR-2002 (first entry)

DE Sheep alpha-S2 casein precursor protein.

XX Sheep; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
KM chewing gum; cosmetic; wrinkling; periodontal disease.

OS Ovis sp.

PN WO200202133-A2.

XX 10-JAN-2002.

XX 13-JUN-2001; 2001WO-GB02601.

XX 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD.

XX Smith JA;

XX WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in
PT alpha-S2 casein precursor in the manufacture of a medicament for
PT alleviating or preventing periodontal disease and an effect of aging in
PT skin -

PS Claim 8; Page 9; 27pp; English.

CC The invention relates to a composition comprising a peptide or its
CC derivative. The peptide contains an amino acid sequence from alpha-S2
CC casein precursor. The peptides stimulate the growth of fibroblasts, and
CC thus the synthesis and secretion of collagen. The peptides also stimulate
CC the growth of keratinocytes, which aid in the formation and regeneration
CC of skin surface. The peptide is useful in the manufacture of a medicament
CC in the form of a toothpaste or a chewing gum, for alleviating or
CC preventing periodontal disease and a medicament in the form of a cosmetic
CC composition for alleviating or preventing an effect of aging,
CC particularly wrinkling of the skin. The present sequence is sheep
CC alpha-S2 casein precursor protein.

XX Sequence 223 AA;

Query Match 82.2%; Score 111; DB 23; Length 223;
Best Local Similarity 83.3%; Pred. No. 2.5e-08;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQXAMKPMWIOPTKVIPYRY 24
DB 199 VDOHQXAMKPMWIOPTKVTAIPYRY 222

RESULT 15

AAE17475
ID AAE17475 standard; Protein: 223 AA.

AC AAE17475;

XX 22-APR-2002 (first entry)

XX Alpha-S2 casein precursor (alpha-S2-CN) protein #2.

DE Alpha-S2 casein; alpha-S2-CN; dermatological; antiinflammatory; cosmetic;
KW fibroblast; collagen; keratinocyte; skin regeneration; medicament; aging;
KM toothpaste; chewing gum; cosmetic; wrinkling; periodontal disease.

XX Unidentified.

OS WO200202133-A2.

XX 10-JAN-2002.

XX 13-JUN-2001; 2001WO-GB02601.

XX 30-JUN-2000; 2000GB-0016189.

XX (PEPS-) PEPSYN LTD.

XX Smith JA;

XX WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in
PT alpha-S2 casein precursor in the manufacture of a medicament for
PT alleviating or preventing periodontal disease and an effect of aging in
PT skin -
PS Disclosure; Page 9; 27pp; English.

CC The invention relates to a composition comprising a peptide or its
CC derivative. The peptide contains an amino acid sequence from alpha-S2
CC casein precursor. The peptides stimulate the growth of fibroblasts, and
CC thus the synthesis and secretion of collagen. The peptides also stimulate
CC the growth of keratinocytes, which aid in the formation and regeneration
CC of skin surface. The peptide is useful in the manufacture of a medicament
CC in the form of a toothpaste or a chewing gum, for alleviating or
CC preventing periodontal disease and a medicament in the form of a cosmetic
CC composition for alleviating or preventing an effect of aging,
CC particularly wrinkling of the skin. The present sequence is alpha-S2
CC casein precursor (alpha-S2-CN) protein.

XX Sequence 223 AA;

Query Match 82.2%; Score 111; DB 23; Length 223;
Best Local Similarity 83.3%; Pred. No. 2.5e-08;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQXAMKPMWIOPTKVIPYRY 24
DB 199 VDOHQXAMKPMWIOPTKVTAIPYRY 222

Search completed: July 30, 2003, 16:23:27
Job time : 48.2857 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 42.4107 Seconds

(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-4

Sequence: 1 VYOHQKMKPWIOPTKTKVLPYRYL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

SPREMBL_23:*
1: sp archaea:*
2: sp bacteria:*
3: sp fungi:*
4: sp human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteria:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	123	88.5	210	6	062825
2	115	82.7	223	6	09MTU6
3	106	76.3	223	6	09TUT7
4	106	76.3	223	6	09TUT7
5	53	38.1	935	3	09GK07
6	52	37.4	80	10	081534
7	52	37.4	85	10	081399
8	52	37.4	340	10	096568
9	52	37.4	379	10	004065
10	52	37.4	386	10	023923
11	52	37.4	388	10	09FS37
12	52	37.4	389	10	08LP19
13	52	37.4	389	10	043040
14	52	37.4	389	10	093XP8
15	52	37.4	389	10	09MSB2
16	52	37.4	390	10	09LBR2

17	52	37.4	390	10	048564	048564 scutellaria
18	52	37.4	390	10	022122	022122 scutellaria
19	51	36.7	394	10	093V86	093V86 humulus lup
20	51	36.7	399	10	094LW8	094LW8 humulus lup
21	50.5	36.3	471	5	015918	015918 plasmodium
22	50	36.0	311	5	019257	019257 caenorhabdi
23	50	36.0	693	11	09WV66	09WV66 mus musculi
24	50	36.0	1877	3	09USR3	09USR3 echinosach
25	49.5	35.6	325	16	09K947	09K947 bacillus ha
26	49.5	35.3	282	5	044074	044074 ascaris suu
27	49	35.3	342	16	08YV06	08YV06 anabaena sp
28	48.5	34.9	286	10	08W2F2	08W2F2 arabidopsis
29	48	34.5	183	16	09CP51	09CP51 pasteurella
30	48	34.5	208	16	08F8G2	08F8G2 leptospira
31	48	34.5	225	4	09BTR9	09BTR9 homo sapien
32	48	34.5	328	16	08CP60	08CP60 staphylococ
33	48	34.5	374	10	081476	081476 brasica na
34	48	34.5	392	10	093YX5	093YX5 vitis sp. c
35	48	34.5	392	10	09PEK7	09PEK7 vitis labru
36	48	34.5	392	10	08LP44	08LP44 parthenocis
37	48	34.5	392	10	09FRW2	09FRW2 vitis ripar
38	48	34.5	392	10	08LPW2	08LPW2 parthenocis
39	48	34.5	392	10	08LPW2	08LPW2 parthenocis
40	48	34.5	392	10	09S982	09S982 vitis. still
41	48	34.5	392	10	08LPW3	08LPW3 cistus rhom
42	48	34.5	392	10	094G58	094G58 vitis vinif
43	48	34.5	392	10	0944W8	0944W8 vitis sp. c
44	48	34.5	392	10	0944W7	0944W7 vitis sp. c
45	48	34.5	395	10	09FR70	09FR70 alliarica pe

ALIGNMENTS

RESULT 1	ID	62825	PREDIMINARY;	PRT;	210 AA.
AC	062825;				
DT	01-AUG-1998 (TREMREL. 07, Created)				
DT	01-NOV-1999 (TREMREL. 12, Last sequence update)				
DT	01-JUN-2001 (TREMREL. 17, Last annotation update)				
DE	AS2-casein (fragment).				
OS	Bubalus bubalis (Domestic water buffalo).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bubalus.				
OX	NCBI_TaxID=89462;				
RN	[1]				
RP	SEQUENCE FROM N. A.				
RC	Tissue=Mammary gland;				
KA	Dae P., Dain S., Garg L.C.;				
RT	"Cloning and nucleotide sequence of cDNA encoding as2-casein in B.				
RT	bubalis."				
RL	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AJ005431; CA06534.2; -				
DR	InterPro; IPR001588; Casein.				
DR	Pfam; PF00363; caseins; 2.				
FT	NON TER				
FT	SEQUENCE 210 AA; 24700 MW; 05DEF95963F1132C CRC64;				
QY	Query Match				
QY	Best Local Similarity 88.5%; Score 123; DB 6; Length 210;				
QY	Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;				
DB	1 VYOHQKMKPWIOPTKTKVLPYRYL 25				
DB	186 VYOHQKMKPWIOPTKTKVLPYRYL 210				
RESULT 2					
QY	09MTU6				
AC	09MTU6; PRELIMINARY; PRT; 223 AA.				

```

DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Alpha 82-casein.
GN CSNIS2.
OC Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OK NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA Lagonigro R., Pilla F., Mataesino D., Zullo A.;
RT "A new allele of goat alpha 82-casein gene."
RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ289716; CAB94236.1; -.
DR InterPro: IPR001588; Casein.
DR Pfam: PF00363; caseins; 2.
DR PROSITE: PS00306; CASEIN_ALPHA_BETA; 1.
SQ SEQUENCE 223 AA; 26403 MW; 0B1F83F24DA85E2 CRC64;

Query Match
Best Local Similarity 82.7%; Score 115; DB 6; Length 223;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHKAMKPMIOPTKVIPIVRYL 25
DB 199 VDQHKAMKPMIOPTKVIPIVRYL 223

RESULT 3
Q9TT07 PRELIMINARY; PRT; 223 AA.
AC Q9TT07;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE Alpha 82-casein.
GN CSNIS2.
OC Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OK NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA Veltre C.C., Pilla F.F., Lagonigro R.R.;
RT "A new allele of goat alpha 82-casein."
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ249995; CAB59920.1; -.
DR InterPro: IPR001588; Casein.
DR Pfam: PF00363; caseins; 2.
DR PROSITE: PS00306; CASEIN_ALPHA_BETA; 1.
SQ SEQUENCE 223 AA; 26433 MW; CE9F4DC8D7688293 CRC64;

Query Match
Best Local Similarity 76.3%; Score 106; DB 6; Length 223;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 VYOHKAMKPMIOPTKVIPIVRYL 25
DB 199 VDQHKAMKPMIOPTKVIPIVRYL 223

RESULT 4
Q9GK07 PRELIMINARY; PRT; 223 AA.
AC Q9GK07;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
DE Alpha 82-casein.
GN CSNIS2.

```

```

OC Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OK NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA Veltre C.C., Pilla F.F., Lagonigro R.R.;
RT "A new allele of alpha 82-casein."
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ297310; CAC21704.2; JOINED.
DR EMBL; AJ297311; CAC21704.2; JOINED.
DR EMBL; AJ242728; CAC21704.2; JOINED.
DR EMBL; AJ297312; CAC21704.2; JOINED.
DR EMBL; AJ297313; CAC21704.2; JOINED.
DR EMBL; AJ297314; CAC21704.2; JOINED.
DR EMBL; AJ297315; CAC21704.2; JOINED.
DR EMBL; AJ297316; CAC21704.2; JOINED.
DR EMBL; AJ242526; CAC21704.2; JOINED.
DR EMBL; AJ242528; CAC21704.2; JOINED.
DR EMBL; AJ242533; CAC21704.2; JOINED.
DR InterPro: IPR001588; Casein.
DR Pfam: PF00363; caseins; 2.
DR PROSITE: PS00306; CASEIN_ALPHA_BETA; 1.
SQ SEQUENCE 223 AA; 26432 MW; CE97658D7688C9D CRC64;

Query Match
Best Local Similarity 76.3%; Score 106; DB 6; Length 223;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 VYOHKAMKPMIOPTKVIPIVRYL 25
DB 199 VDQHKAMKPMIOPTKVIPIVRYL 223

RESULT 5
O94623 PRELIMINARY; PRT; 935 AA.
ID O94623;
AC O94623;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Hypothetical 106.5 kDa protein C1347.01C in chromosome II.
GN SPBC1347.01C OR SPBC215.16C.
OC Schizosaccharomyces pombe (Pisition yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Rieger M.;
RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 731-935 FROM N.A.
RC STRAIN=972;
RA Lyne M., Rajandream M.A., Barrell B.G., Rieger M.;
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: TO C. ELBEGANS ZK675.2.
CC -1- SIMILARITY: BELONGS TO THE UMC (DNA REPAIR) FAMILY.
DR EMBL; AL035548; CAB37432.1; -.
DR EMBL; AL033534; CAB22130.1; -.
DR HSRP; P96022; IIMA.
DR GeneDB; sPombe; SPBC1347.01C; -.
DR InterPro: IPR001357; BRCT.
DR InterPro: IPR001126; UMC_1like.
DR Pfam; PF00533; BRCT; 1.
DR Pfam; PF00817; IMS; 1.
DR SMART; SMO0292; BRCT; 1.

```


RA Gong Z., Yamazaki M., Sugiyama M., Tanaka Y., Saito K.;
 RT "Cloning and molecular analysis of structural genes involved in
 RT anthocyanin biosynthesis and expressed in a forma-specific manner in
 RT *Pteris caudata*.";
 RL Plant Mol. Biol. 35:915-927(1997).
 DR EMBL; AB002582; BAA19548.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_still_synth; 1.
 DR Pfam; PF02797; Chal_still_synthC; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 379 AA; 4180 MW; 5E149C55530BD106 CRC64;

QY Query Match 37.4%; Score 52; DB 10; Length 379;
 Best Local Similarity 64.3%; Pred. No. 9.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 5 OKAMKPMIOPKTV 18
 111 OKAIKMGQPKSKI 124

RESULT 10
 ID 023923 PRELIMINARY; PRT; 386 AA.
 AC 023923;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Chalcone synthase (Fragment).
 OS Digitalis lanata (Foxglove).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Lamiales; Antirrhinaceae; Digitalideae; Digitalis.
 OK NCBI_TaxID=49450;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=leaf;
 RA Theoringer C., Lindemann P., Luckner M.;
 RT "Expression of phenylalanine ammonia lyase, chalcone synthase and
 RT phytochrome in somatic embryogenesis of *Digitalis lanata*.";
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ002526; CAA05512.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_still_synth; 1.
 DR Pfam; PF02797; Chal_still_synthC; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 FT NON_TER 1
 SQ SEQUENCE 386 AA; 42190 MW; 0A88B8FED828F8 CRC64;

QY Query Match 37.4%; Score 52; DB 10; Length 386;
 Best Local Similarity 64.3%; Pred. No. 10;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 5 OKAMKPMIOPKTV 18
 106 OKAIKMGQPKSKI 119

RESULT 11
 ID 09F637 PRELIMINARY; PRT; 388 AA.
 AC 09F637;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Chalcone synthase.
 OS *Torenia hybrida*.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; lamids; Lamiales; Lamiales incertae sedis; Toreneae;
 OC *Torenia*.
 OK NCBI_TaxID=75807;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Summerwave Blue; TISSUE=Petal;
 RC Suzuki K., Xue H., Tanaka Y., Fukui Y., Fukushima-Mizutani M.,
 RA Murakami Y., Katsumoto Y., Tsuda S., Kusumi T.;
 RT "Flower color modifications of *Torenia hybrida* by cosuppression of
 RT anthocyanin biosynthesis genes.";
 RL Mol. Breed. 6:239-246(2000).
 DR EMBL; AB012923; BAB20074.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_still_synth; 1.
 DR Pfam; PF02797; Chal_still_synthC; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 388 AA; 42402 MW; BDD3A78B6C44A8B0F CRC64;

QY Query Match 37.4%; Score 52; DB 10; Length 388;
 Best Local Similarity 64.3%; Pred. No. 10;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 5 OKAMKPMIOPKTV 18
 111 OKAIKMGQPKSKI 124

RESULT 12
 ID 08LP19 PRELIMINARY; PRT; 389 AA.
 AC 08LP19;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Chalcone synthase.
 GN CHS.
 OS *Nierembergia* sp. NB17.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Nierembergia.
 OK NCBI_TaxID=184877;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NB17;
 RA Ueyama Y., Katsumoto Y., Fukui Y., Ohkawa H., Kusumi T., Tanaka Y.;
 RT "Flower color modification of *Nierembergia* sp. by engineering
 RT flavonoid biosynthetic pathway.";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB078515; BAC10998.1; -.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_still_synth; 1.
 DR Pfam; PF02797; Chal_still_synthC; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 389 AA; 42616 MW; 62DA0CTF40729E7 CRC64;

QY Query Match 37.4%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 10;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 5 OKAMKPMIOPKTV 18
 111 OKAIKMGQPKSKI 124

RESULT 13
 ID 043040 PRELIMINARY; PRT; 389 AA.
 AC 043040;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Chalcone synthase.
 GN CHS-A
 OS Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Petunia.
 NCBI_TaxID=4102;
 RX MEDLINE=96156377; PubMed=8562849;
 RA Shao L., Li Y., Pan A., Cheng Z., Chen M.;
 RT "Molecular cloning, sequencing, and expression in *Escherichia coli* of
 the chalcone synthase gene."
 RL Chin. J. Biotechnol. 11:131-135(1995).
 DR EMBL/580857; AAB36038.1; -
 DR HSSP; P30074; ID6F.
 DR InterPro: IPR001099; N-C synthase.
 DR Pfam; PF00195; Chal_stil_synth; 1.
 DR Pfam; PF02797; Chal_stil_synth; 1.
 DR Prodom; PD000453; N-C synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 389 AA; 42642 MW; A93CF5B255A0A20 CRC64;
 Query Match 37.4%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 10;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 5 OKAMKPMIOPKTKV 18
 DB 111 OKAIKMGQPKSKI 124
 RESULT 14
 ID Q93XP8 PRELIMINARY; PRT; 389 AA.
 AC Q93XP8;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Chalcone synthase.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
 NCBI_TaxID=4097;
 RX TISSUE=Flower;
 RA Hu H., Sung H., Su J.;
 RT "Cloning and Expression of a Tobacco Chalcone Synthase Gene."
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF311783; AAK9457.1; -
 DR InterPro: IPR001099; N-C synthase.
 DR Pfam; PF00195; Chal_stil_synth; 1.
 DR Pfam; PF02797; Chal_stil_synth; 1.
 DR Prodom; PD000453; N-C synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 389 AA; 42563 MW; 96C3D426706BBD44 CRC64;
 Query Match 37.4%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 10;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 5 OKAMKPMIOPKTKV 18
 DB 111 OKAIKMGQPKSKI 124
 RESULT 15
 Q9MSB2 PRELIMINARY; PRT; 389 AA.
 AC Q9MSB2;

DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Chalcone synthase (EC 2.3.1.74).
 GN CHS.
 OS Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Petunia.
 NCBI_TaxID=4102;
 RX [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Violett; TISSUE=Corolla;
 RA Hsu Y.H., Durdan S.F., To K.Y.;
 RT "Cloning of a pigment-related cDNA encoding chalcone synthase,
 RT chalcone isomerase, dihydroflavonol-4-reductase and cytochrome b5 from
 RT Petunia hybrida."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF233638; AAF60297.1; -
 DR HSSP; P30074; ID6F.
 DR InterPro: IPR001099; N-C synthase.
 DR Pfam; PF00195; Chal_stil_synth; 1.
 DR Pfam; PF02797; Chal_stil_synth; 1.
 DR Prodom; PD000453; N-C synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 DR Acyltransferase; Transferase.
 SQ SEQUENCE 389 AA; 42582 MW; 66135700259D7F0C CRC64;
 Query Match 37.4%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 10;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 5 OKAMKPMIOPKTKV 18
 DB 111 OKAIKMGQPKSKI 124
 Search completed: July 30, 2003, 16:29:23
 Job time : 43.4107 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:10:10 ; Search time 8.18452 Seconds
(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-4
Sequence score: 139
Sequence: 1 VYOHQRMKPMWIOPKVKVLPYVRYL 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match	Length	ID	Description
1	139	100.0	222	1 CAS2_BOVIN	P02663 bos taurus
2	115	82.7	223	1 CAS2_CAPRI	P33049 capra hircu
3	115	82.7	223	1 CAS2_SHEEP	P04654 ovis aries
4	70.5	50.7	225	1 CAS2_PIG	P39036 sus scrofa
5	67	48.2	183	1 CAS2_RABIT	P50419 oryctolagus
6	55	39.6	192	1 CAS2_CAMDR	O97944 camelus dro
7	52	37.4	389	1 CHS1_LYCES	P23418 lycopersico
8	52	37.4	389	1 CHS2_LYCES	P23418 lycopersico
9	52	37.4	389	1 CHS2_SOLITU	P08899 solanum tub
10	52	37.4	389	1 CHS2_PETHY	P08899 petunia hyb
11	52	37.4	389	1 CHS2_SOLITU	Q41436 solanum tub
12	52	37.4	389	1 CHS2_SOLITU	Q41436 solanum tub
13	52	37.4	389	1 CHS2_IPOPU	O21047 ipomoea pur
14	52	37.4	389	1 CHS2_PPHANT	O22046 pharbitis n
15	52	37.4	389	1 CHS2_PETHY	P22928 petunia hyb
16	52	37.4	390	1 CHS2_PETHY	P06515 antirrhinum
17	52	37.4	391	1 CHS2_PETHY	O04111 petilla fru
18	52	37.4	400	1 CHS2_PETHY	O04111 petilla fru
19	52	37.4	401	1 CHS2_SORBI	Q09431 sorghum bic
20	52	37.4	401	1 CHS2_SORBI	Q09431 sorghum bic
21	52	37.4	401	1 CHS2_SORBI	Q09431 sorghum bic
22	52	37.4	401	1 CHS2_SORBI	Q09431 sorghum bic
23	52	37.4	401	1 CHS2_SORBI	Q09431 sorghum bic
24	52	37.4	401	1 CHS2_SORBI	Q09431 sorghum bic
25	50	36.0	291	1 CAS2_RABIT	P50418 oryctolagus
26	49.5	35.6	564	1 YSG6_STRUC	O91159 mus muscul
27	49	35.3	207	1 YSG6_STRUC	O91159 mus muscul
28	49	35.3	382	1 CHS2_SECC	P44766 mus muscul
29	49	35.3	382	1 CHS2_SECC	P44766 mus muscul
30	49	35.3	382	1 CHS2_SECC	P44766 mus muscul
31	49	35.3	382	1 CHS2_SECC	P44766 mus muscul
32	48	34.5	389	1 THS3_VITVI	P23343 vitis vinif
33	48	34.5	392	1 THS3_VITVI	P23343 vitis vinif

34	48	34.5	392	1 THS2_VITVI	P51070 vitis vinif
35	48	34.5	5171	1 BPSA_HUMAN	O04833 homo sapien
36	47	33.8	285	1 CHS6_MEDSA	P51079 medicago sa
37	47	33.8	311	1 CHS4_TIRISU	P51086 trifolium s
38	47	33.8	388	1 CHS1_SOYEN	P24826 glycine max
39	47	33.8	388	1 CHS2_SOYEN	P19587 glycine max
40	47	33.8	388	1 CHS3_SOYEN	P19168 glycine max
41	47	33.8	388	1 CHS5_SOYEN	P48406 glycine max
42	47	33.8	388	1 CHS6_SOYEN	P30080 glycine max
43	47	33.8	388	1 CHS8_TROBA	Q0636 ipomoea bat
44	47	33.8	388	1 CHSD_PPHANT	O22045 pharbitis n
45	47	33.8	388	1 CHS_VITGUN	P51089 vigna ungu

ALIGNMENTS

RESULT 1
CAS2_BOVIN STANDARD; PRT; 222 AA.
AC P02663; Q9TR51;
DT 21-JUL-1986 (Ref. 01, Created)
DT 01-JUL-1989 (Ref. 11, Last sequence update)
DT 28-FEB-2003 (Ref. 41, Last annotation update)
DE Alpha-S2 casein precursor [contains: Casocidin-I].
GN CSH1S2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_Taxid=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88188989; PubMed=2833669;
RA Stewart A.F., Bonsting J., Beattie C.W., Shah F., Willis I.M.,
RA Mackinlay A.G.;
RT "Complete nucleotide sequences of bovine alpha S2- and beta-casein
RT cDNAs: comparisons with related sequences in other species.";
RT Mol. Biol. Evol. 4:231-241(1987).
RN [2]
RP SEQUENCE OF 16-222 (A ALLELE).
RX TISSUE=Milk;
RA MEDLINE=77185633; PubMed=862906;
RA Brignon G., Ribadeau-Dumas B., Mercier J.-C., Pellissier J.-P.,
RA Das B.C.;
RT "Complete amino acid sequence of bovine alpha2-casein.";
RT FEBS Lett. 76:274-279(1977).
RN [3]
RP PARTIAL SEQUENCE (D ALLELE).
RX TISSUE=Milk;
RA MEDLINE=79239837; PubMed=469044;
RA Grosclaude F., Joudrier P., Mahe M.-F.;
RT "A genetic and biochemical analysis of a polymorphism of bovine alpha
RT S2-casein.";
RT J. Dairy Res. 46:211-213(1979).
RN [4]
RP SEQUENCE OF 165-203, AND CHARACTERIZATION OF CASOCIDIN.
RX TISSUE=Milk;
RA MEDLINE=9600204; PubMed=7556666;
RA Zucht H.-D., Raída M., Adernann K., Meagert H.-J., Forssmann W.-G.;
RT "Casocidin-I: a casein-alpha s2 derived peptide exhibits antibacterial
RT activity.";
RT FEBS Lett. 372:185-188(1995).
CC - FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC - FUNCTION: CASOCIDIN-I INHIBITS THE GROWTH OF ESCHERICHIA COLI AND
CC STREPTOCOCCUS CARNOBACTER.
CC - SUBCELLULAR LOCATION: Extracellular.
CC - TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC - MASS SPECTROMETRY: MW=4870; METHOD=Electrospray; RANGE=165-203.
CC - POLYMORPHISM: AT LEAST TWO ALLELES EXIST. THE SEQUENCE OF THE A
CC ALLELE IS SHOWN HERE. THE D ALLELE SEQUENCE DIFFERS FROM THAT
CC SHOWN IN HAVING A DELETION OF NINE RESIDUES, WHICH MAY BE 49-58,

```

CC 50-59, OR 51-60.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -1- DATABASE: NAME=Protein Spotlight;
CC NOTE=Issue 16 of November 2001;
CC WWW="http://www.expasy.org/spotlight/articles/sp11016.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M16644; AAA30479.1; -.
CC PIR; J02008; KABOS2.
CC InterPro; IPR001588; Casein.
CC Pfam; PF00363; caseins; 2.
CC PROSITE; PS00306; CASEIN ALPHA BETA; 1.
CC Milk; Phosphorylation; Signal; Repeat; Antibiotic.
CC SIGNAL 1 15
CC CHAIN 1 15
CC PEPTIDE 16 222 ALPHA-S2 CASEIN.
CC REPEAT 76 203 CASOCIDIN-1.
CC REPEAT 158 222
CC MOD RES 23 23 PHOSPHORYLATION.
CC MOD RES 24 24 PHOSPHORYLATION.
CC MOD RES 25 25 PHOSPHORYLATION.
CC MOD RES 31 31 PHOSPHORYLATION.
CC MOD RES 71 71 PHOSPHORYLATION.
CC MOD RES 72 72 PHOSPHORYLATION.
CC MOD RES 73 73 PHOSPHORYLATION.
CC MOD RES 76 76 PHOSPHORYLATION.
CC MOD RES 144 144 PHOSPHORYLATION.
CC MOD RES 146 146 PHOSPHORYLATION.
CC CONFLICT 102 102 Q -> E (IN REF. 2).
CC FT MOD RES 102 102
CC FT CONFLICT 102 102
CC SQ SEQUENCE 222 AA; 26019 MW; 81E7408AF6C12F7C CRC64;

Query Match 100.0%; Score 139; DB 1; Length 222;
Best Local Similarity 100.0%; Pred. No. 2.9e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMWIOPTKTVIPYRYL 25
Db 198 VYOHOKAMKPMWIOPTKTVIPYRYL 222

RESULT 2
CAS2_CAPHI STANDARD; PRT; 223 AA.
ID CAS2_CAPHI
AC P13079;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-S2 casein precursor (Alpha-S2-CN).
GN CSNIS2.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OK NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93216130; PubMed=846280;
RA Bouniol C.;
RT "Sequence of the goat alpha s2-casein-encoding cDNA."
RL Gene 125:235-236(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX Bouniol C.; Brignon G.; Mahe M.-F.; Prinz C.;
RT "Characterization of goat allelic alpha-s2-caseins A and B: further
RT evidence of the phosphorylation code of caseins."
RL Protein Seq. Data Anal. 5:213-218(1993).

```

```

RN [3]
RP SEQUENCE FROM N.A. (VARIANT C).
RX MEDLINE=95030556; PubMed=7943951;
RA Bouniol C.; Brignon G.; Mahe M.-F.; Prinz C.;
RT "Biochemical and genetic analysis of variant C of caprine alpha s2-
RT casein (Capra hircus)."
RL Ann. Genet. 25:173-177(1994).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- POLYMORPHISM: THREE ALLELES OF ALPHA-S2 CASEIN ARE KNOWN: A, B AND
CC C. THE FREQUENCIES OF THE ALLELES IS ESTIMATED TO BE 0.85, 0.04
CC AND 0.11 IN THE FRENCH DAIRY BREEDS 'ALPINE' AND 'SAENEN'.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X65160; CAA46278.1; -.
CC PIR; S74171; AAB32166.1; -.
CC PIR; I46995; I46995.
CC PIR; S33881; JN0547.
CC InterPro; IPR001588; Casein.
CC Pfam; PF00363; caseins; 2.
CC PROSITE; PS00306; CASEIN ALPHA BETA; 1.
CC Milk; Phosphorylation; Signal; Repeat; Polymorphism.
CC SIGNAL 1 15
CC CHAIN 1 15
CC PEPTIDE 16 223 ALPHA-S2 CASEIN.
CC REPEAT 77 141
CC REPEAT 159 223
CC MOD RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
CC MOD RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
CC MOD RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
CC MOD RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
CC MOD RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
CC MOD RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
CC MOD RES 77 77 PHOSPHORYLATION (BY SIMILARITY).
CC MOD RES 145 145 PHOSPHORYLATION (BY SIMILARITY).
CC MOD RES 147 147 PHOSPHORYLATION (BY SIMILARITY).
CC VARIANT 79 79 E -> K (IN VARIANT B).
CC VARIANT 182 182 K -> I (IN VARIANT C).
CC FT VARIANT 182 182
CC FT SEQUENCE 223 AA; 26389 MW; 187DEF42FD688291 CRC64;

Query Match 82.7%; Score 115; DB 1; Length 223;
Best Local Similarity 84.0%; Pred. No. 9.6e-10;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMWIOPTKTVIPYRYL 25
Db 199 VDOHOKAMKPMWIOPTKTVIPYRYL 223

RESULT 3
CAS2_SHEEP STANDARD; PRT; 223 AA.
ID CAS2_SHEEP
AC P04654;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-S2 casein precursor.
GN CSNIS2.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OK NCBI_TaxID=9940;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=86104467; PubMed=3002499;
RA Boissard M., Petrican G.;
RT "Complete sequence of ovine alpha s2-casein messenger RNA.";
RL Biochimie 67:1043-1051(1985).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X03238; CA26983.1; -.
DR PIR; A25070; KASHS2.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN ALPHA BETA; 1.
KW Milk; Phosphorylation; Repeat; Signal.
FT SIGNAL 1 15
FT CHAIN 1 15
FT REPEAT 16 223 ALPHA-S2 CASEIN.
FT REPEAT 77 141
FT REPEAT 159 223
FT MOD_RES 23 23 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 24 24 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 25 25 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 72 72 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 73 73 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 74 74 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 74 74 PHOSPHORYLATION (POTENTIAL).
FT VARIANT 64 64 D -> N.
SQ SEQUENCE 223 AA; 26332 MW; 67212935E27426D7 CRC64;

Query Match 82.7%; Score 115; DB 1; Length 223;
Best Local Similarity 84.0%; Pred. No. 9.6e-10;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKMKRPMIOPTKVIPIRYRYL 25
Db 199 VDOHQKMKRPMIOPTKVIPIRYRYL 223

RESULT 4
ID CAS2_PIG STANDARD; PRT; 235 AA.
AC P39036;
RT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DB Alpha-S2 casein precursor.
GN CSNIS2.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=99823;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Mammary gland;
RX MEDLINE=92367960; PubMed=1503276;
RA Alexander L.J., Das Gupta N.A., Beattie C.W.;
RT "The sequence of porcine alpha s2-casein cDNA.";
RL Annu. Genet. 23:365-367(1992).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
-----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X54975; CA38719.1; -.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN ALPHA BETA; 1.
KW Milk; Phosphorylation; Signal.
FT SIGNAL 1 15
FT CHAIN 1 15
FT MOD_RES 16 235 ALPHA-S2 CASEIN.
FT MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 70 70 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 71 71 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 235 AA; 27570 MW; C903B760D184C14C CRC64;

Query Match 50.7%; Score 70.5; DB 1; Length 235;
Best Local Similarity 55.6%; Pred. No. 0.0034;
Matches 15; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

QY 1 VYOHQKMKRPMIOPTKVIPIRYRY 24
Db 208 VDOHQKMKRPMIOPTKVIPIRYRY 234

RESULT 5
ID CAS3_RABIT STANDARD; PRT; 182 AA.
AC P50419;
RT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DB Alpha-S2B casein precursor.
GN Oryctolagus cuniculus (Rabbit).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=New Zealand white; TISSUE=Mammary gland;
RX MEDLINE=94107245; PubMed=8280077;
RX Dawson S.P., Wilde C.J., Tighe P.J., Mayer R.U.;
RT "Characterization of two novel casein transcripts in rabbit mammary
RT gland.";
RL Biochem. J. 296:777-784(1993).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X76909; CA54231.1; -.
DR PIR; S39776; S39776.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.

```

DR PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
 KM Milk; Phosphorylation; Signal.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 16 182 ALPHA-S2B CASEIN.
 SQ SEQUENCE 182 AA; 21597 MW; F2433C2DA45550FC CRC64;
 Query Match 48.2%; Score 67; DB 1; Length 182;
 Best Local Similarity 43.5%; Pred. No. 0.0085;
 Matches 10; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 VYOHAKKMPWIOKTYIPYV 23
 DB 83 LYHPTVMDPWTAKYTAIPR 105

RESULT 6
 CAS2_CAMDR STANDARD; PRT; 193 AA.
 ID CAS2_CAMDR STANDARD; PRT; 193 AA.
 AC 097844;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Alpha-S2 casein precursor.
 GN CSNIS2.
 OS Camelus dromedarius (Dromedary) (Arabian camel).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
 ON NCBI_Taxid=9818;
 RX [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN=Somali; TISSUE=Udder;
 RX MEDLINE=98291310; PubMed=9627840;
 RA Kappeler S., Farah Z., Puhaz Z.;
 RT "Sequence analysis of Camelus dromedarius milk caseins."
 RL J. Dairy Res. 65:209-222(1998).
 CC - FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
 CC - CALCIUM PHOSPHATE.
 CC - SUBCELLULAR LOCATION: Extracellular.
 CC - TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
 CC - SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC - This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ012629; CAA10078.1; -.
 DR InterPro: IPR001588; Casein.
 DR Pfam: PF00363; caseins; 2.
 DR PROSITE; PS00306; CASEIN_ALPHA_BETA; FALSE_NRG.
 KM Milk; Phosphorylation; Signal.
 FT SIGNAL 1 15
 FT CHAIN 16 193 ALPHA-S2 CASEIN.
 SQ SEQUENCE 193 AA; 22964 MW; 2843356F8F2ED13 CRC64;

Query Match 39.6%; Score 55; DB 1; Length 193;
 Best Local Similarity 40.9%; Pred. No. 0.52;
 Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 VYOHAKKMPWIOKTYIPYV 22
 DB 93 LHOGIVNPNWDQCKTRAYPFI 114

RESULT 7
 CHS1_LYCSES STANDARD; PRT; 389 AA.
 ID CHS1_LYCSES STANDARD; PRT; 389 AA.
 AC P23418;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Chalcone synthase 1 (EC 2.3.1.74) (Naringenin-chalcone synthase 1).
 GN CHS1.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Solanaceae; Solanum.
 ON NCBI_Taxid=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=cotyledon, hypocotyl and leaf;
 RX MEDLINE=91117196; PubMed=1980524;
 RA O'Neill S.D., Tong Y., Spoerlein B., Forkmann G., Yoder J.I.;
 RT "Molecular genetic analysis of chalcone synthase in Lycopersicon
 RT esculentum and an anthocyanin-deficient mutant."
 RL Mol. Gen. Genet. 224:279-288(1990).
 CC - FUNCTION: The primary product of this enzyme is 4',2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC - CATALYTIC ACTIVITY: 3 malonyl-CoA + 4 coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC - PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC - SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC - This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X55194; CAA38980.1; -.
 DR InterPro: IPR001099; N-C-synthase.
 DR Pfam: PF00195; Chal_ec11_synth; 1.
 DR Pfam: PF02797; Chal_ec11_synth; 1.
 DR ProDom: PD000453; N-C synthase; 1.
 DR PROSITE; PS00441; CHALCONE SYNTH; 1.
 KM Flavonoid biosynthesis; Transferase; Acyltransferase;
 KW Multigene family.
 FT ACT SITE 164 164 BY SIMILARITY.
 SQ SEQUENCE 389 AA; 42552 MW; 553DC695EA96A8B CRC64;

Query Match 37.4%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 2.9;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QKAKMPWIOKTYIPYV 18
 DB 111 OKATKMGOPKSKI 124

RESULT 8
 CHS2_LYCSES STANDARD; PRT; 389 AA.
 ID CHS2_LYCSES STANDARD; PRT; 389 AA.
 AC P23419;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2).
 GN CHS2.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Solanaceae; Solanum.
 ON NCBI_Taxid=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cotyledon, Hypocotyl, and leaf;
 RX MEDLINE=91117196; PubMed=1980524;

RA O'Neill S.D., Tong Y., Speerlein B., Forkmann G., Yoder J.I.;
 RT "Molecular genetic analysis of chalcone synthase in lycopersicon
 RL esculentum and an anthocyanin-deficient mutant.";
 CC Mol. Genet. 224:279-288(1990).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X55195; CAA38981.1; -;
 CC InterPro: IPR001099; N-C synthase.
 CC Pfam: PF00195; Chal_stil_synth; 1.
 CC DR Pfam: PF02797; Chal_stil_synth; 1.
 CC DR PRODOM: PD000453; N-C_synthase; 1.
 CC DR PROSITE: PS00441; CHALCONE_SYNTH; 1.
 CC KM Flavonoid biosynthesis; Transferase; Acyltransferase;
 CC KW Multigene family.
 CC FT ACT SITE 164 164
 CC FT ACT SITE 164 164
 CC BY SIMILARITY.
 CC SEQUENCE 389 AA; 42730 MW; F92E46BB3B5FC32F CRC64;
 CC
 CC Query Match 37.4%; Score 52; DB 1; Length 389;
 CC Best Local Similarity 64.3%; Pred. No. 2.9;
 CC Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC QY 5 QKAKKPMWIOPTKV 18
 CC |||||
 CC Db 111 QKAIKMGQPKSKI 124
 CC
 CC RESULT 9
 CC CHS2_SOLTU STANDARD; PRT; 389 AA.
 CC AC Q43188;
 CC DT 15-JUL-1999 (Rel. 38, Created)
 CC DT 15-JUL-1999 (Rel. 38, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DB Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2).
 CC GN CHS2.
 CC OS Solanum tuberosum (potato).
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
 CC OC NCBI_TaxID=4113;
 CC RN (1)
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=cv. Red Pontiac;
 CC RX MEDLINE=97141614; PubMed=987872;
 CC RA Jeon J.-H., Kim H.S., Choi K.H., Jung Y.H., Jung H., Byun S.-M.;
 CC RT "Cloning and characterization of one member of the chalcone synthase
 CC gene family from Solanum tuberosum L.";
 CC RL Biochem. Biotechnol. Biochem. 60:1907-1910(1996).
 CC CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY

CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U47738; AAB05239.1; -;
 CC DR PIR: JCS136; JCS136.
 CC DR InterPro: IPR001099; N-C synthase.
 CC DR Pfam: PF00195; Chal_stil_synth; 1.
 CC DR Pfam: PF02797; Chal_stil_synth; 1.
 CC DR PRODOM: PD000453; N-C synthase; 1.
 CC DR PROSITE: PS00441; CHALCONE_SYNTH; 1.
 CC KM Flavonoid biosynthesis; Transferase; Acyltransferase;
 CC KW Multigene family.
 CC FT ACT SITE 164 164
 CC FT ACT SITE 164 164
 CC BY SIMILARITY.
 CC SEQUENCE 389 AA; 42476 MW; 41618F9944958603 CRC64;
 CC
 CC Query Match 37.4%; Score 52; DB 1; Length 389;
 CC Best Local Similarity 64.3%; Pred. No. 2.9;
 CC Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC QY 5 QKAKKPMWIOPTKV 18
 CC |||||
 CC Db 111 QKAIKMGQPKSKI 124
 CC
 CC RESULT 10
 CC CHSA_PETTY STANDARD; PRT; 389 AA.
 CC AC P08894;
 CC DT 01-NOV-1988 (Rel. 09, Created)
 CC DT 01-NOV-1988 (Rel. 09, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Chalcone synthase A (EC 2.3.1.74) (Naringenin-chalcone synthase A).
 CC GN CHSA.
 CC OS Petunia hybrida (Petunia).
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC OC Asteridae; lamids; Solanales; Solanaceae; Petunia.
 CC OC NCBI_TaxID=4102;
 CC RN (1)
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=cv. Violet 30;
 CC RX MEDLINE=86286540; PubMed=3016642;
 CC RA Koes R.E., Spelt C.E., Relf H.U., van den Elzen P.J.M., Veltkamp E.,
 CC RA Mol J.N.M.;
 CC RT "Floral tissue of Petunia hybrida (V30) expresses only one member of
 CC the chalcone synthase multigene family.";
 CC RL Nucleic Acids Res. 14:5229-5239(1986).
 CC RN (2)
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=cv. Violet 30; TISSUE=leaf;
 CC RX MEDLINE=90034197; PubMed=2806915;
 CC RA Koes R.E., Spelt C.E., van den Elzen P.J.M., Mol J.N.M.;
 CC RT "Cloning and molecular characterization of the chalcone synthase
 CC multigene family of Petunia hybrida.";
 CC RL Gene 81:245-257(1989).
 CC CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- TISSUE SPECIFICITY: MAJOR EXPRESSED MEMBER OF THE GENE FAMILY IN

	CC	VARIOUS FLORAL TISSUES AND IN SEEDLINGS TREATED WITH UV LIGHT. IT IS RELATIVELY LOW EXPRESSED IN TISSUE CULTURE MATERIAL.
	CC	-1- SIMILARITY: BELONGS TO THE CHALCONE/S stilbene SYNTHASES FAMILY.
	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licenses@db-sek.ch).
	DR	EMBL; X04080; CAA27718.1; -
	DR	EMBL; X14591; CAA52731.1; -
	DR	PIR; A23643; SYPJCN.
	DR	PIR; J50308; SYPCA.
	DR	InterPro; IPR001039; N-C synthase.
	DR	pfam; PF00195; Chal_stil_synth_1.
	DR	Pfam; PF02797; Chal_stil_synthc_1.
	DR	Prodom; PD000453; N-C synthase; 1.
	DR	PROSITE; PS00441; CHALCONE_SYNTH; 1.
	KW	F flavonoid biosynthesis; transferase; Acyltransferase;
	KM	Multigene family.
	FT	ACT SITE 164 BY SIMILARITY.
	FT	CONFLICT 128 V -> F (IN REF. 2).
	SQ	SEQUENCE 389 AA; 42525 MW; 0221BDDA1E9EA6F CRC64;
	QY	Query Match 37.4%; Score 52; DB 1; Length 389; Best Local Similarity 64.3%; Pred. NO. 2.9; Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0.
	Dz	5 QKAKMKPWIQPKTKV 18 :: : 111 OKAIKMGOPKSRI 124
	RESULT 11	
	CHSA_SOLTU	STANDARD; PRT; 389 AA. ID_CHSA_SOLTU STANDARD; PRT; 389 AA. AC_QAI436; DT 15-JUL-1999 (Rel. 38, Created) DT 15-JUL-1999 (Rel. 38, Last sequence update) DT 28-FEB-2003 (Rel. 41, Last annotation update) DE Chalcone synthase 1A (EC 2.3.1.74) (Naringenin-chalcone synthase 1A). GN CHSLA. OS Solanum tuberosum (Potato). OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; OC Asteridae; lamids; Solanales; Solanaceae; Solanum. OX NCBI_TaxId=4113; RN [1] RP SEQUENCE FROM N.A. RA STRAIN=cv. Red Pontiac; RC Jeon J.-H., Kim H.-S., Choi K.-H., Jung Y.-H., Young H., Byns S.-M.; RL "Characterization of two members of the chalcone synthase gene family from Solanum tuberosum L."; RT (in) Plant Gene Register PG96-027. RL CC -1- FUNCTION: The primary product of this enzyme is 4',2'',4',6'-tetrahydrochalcone (also termed naringenin-chalcone or chalcone) which can under specific conditions spontaneously isomerize into naringenin. CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coannaroyl-CoA = 4 CoA + naringenin-chalcone + 3 CO(2) CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF FLAVONOIDS. A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY OF WHICH ARE BRIGHTLY COLORED. CC -1- SIMILARITY: BELONGS TO THE CHALCONE/stilbene SYNTHASES FAMILY. CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

```
CC   either requires a license agreement (See http://www.isb-sib.ch/announce/  
CC   or send an email to license@isb-sib.ch).  
-----  
CC  
DR   EMBL; U47739; AA867734.1; -.  
DR   InterPro; IPR001099; N-C_synthase.  
DR   Pfam; PF00195; Chal_gtl1_sync; 1.  
DR   Pfam; PF02797; Chal_gtl1_sync; 1.  
DR   ProDom; PD000453; N-C_synthase; 1.  
DR   PROSITE; PS00441; CHALCONE_SYNTH_1.  
KW   Flavonoid biosynthesiss; Transferase; Acyltransferase;  
KW   Multigene family.  
FT   ACT_SITE 164.  
SQ   SEQUENCE 389 AA; 42562 MW; B181D9CBB9170A34 CRC64;  
  
Query Match 37.4%; Score 52; DB 1; Length 389;  
Best Local Similarity 64.3%; Pred. No. 2.9;  
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 5 OKAKMKWIPQKTY 18  
Db |||||:||||:  
111 QAKIKMGQPSKI 124  
  
RESULT 12  
CHSB_SOLUTU BY SIMILARITY  
ID CHSB_SOLUTU STANDARD; PRT; 389 AA.  
AC 043163;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Chalcone synthase 1B (EC 2.3.1.74) (Naringenin-chalcone synthase 1B).  
GN CHS1B.  
OS Solanum tuberosum (potato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.  
OX NCBI_TaxId=4113;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Red Pontiac;  
RA Joon J.-H., Kim H.-S., Choi K.-H., Joung Y.-H., Joung H., Byun S.-M.;  
RT "Characterization of two members of the chalcone synthase gene family  
from Solanum tuberosum L.";  
RL (in) Plant Gene Register PCR96-027.  
CC -!- FUNCTION: The primary product of this enzyme is 4',2'',4'',6''-  
tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)  
which can under specific conditions spontaneously isomerize into  
naringenin.  
CC -!- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +  
naringenin-chalcone + 3 CO(2).  
CC -!- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF  
FLAVONOIDS. A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY  
OF WHICH ARE BRIGHLY COLORED.  
CC -!- SIMILARITY: BELONGS TO THE CHALCONE/SITLBENE SYNTHASIS FAMILY.  
-----  
CC  
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC   the European Bioinformatics institute. There are no restrictions on its  
CC   use by non-profit institutions as long as its content is in no way  
CC   modified and this statement is not removed. Usage by and for commercial  
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC   or send an email to license@isb-sib.ch).  
-----  
CC  
DR   EMBL; U47740; AA867735.1; -.  
DR   InterPro; IPR001099; N-C_synthase.  
DR   Pfam; PF00195; Chal_gtl1_sync; 1.  
DR   Pfam; PF02797; Chal_gtl1_sync; 1.  
DR   ProDom; PD000453; N-C_synthase; 1.  
DR   PROSITE; PS00441; CHALCONE_SYNTH_1.  
KW   Flavonoid biosynthesiss; Transferase; Acyltransferase;  
KW   Multigene family.  
FT   ACT_SITE 164.  
SQ   SEQUENCE 389 AA; 42548 MW; E7334AIN3408BDICC CRC64;  
BY SIMILARITY.
```


Query Match 37.4%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 2.9;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QKAKMPWIOPTKV 18
 |||:|:|:|:|:
 DB 111 OKAIKMGOPKSKI 124

RESULT 13

CHSE_IPOPU STANDARD; PRT; 389 AA.

AC 022047;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chalcone synthase E (EC 2.3.1.74) (Naringenin-chalcone synthase E)
 DE (CHS-E).
 GN CHSE.

OS Ipomoea purpurea (Common morning glory).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Convolvulaceae; Ipomoea.
 NCBI_TaxID=4121;
 [1]

SEQUENCE FROM N.A.

RC STRAIN=CV FR-35; TISSUE=Flower buds;
 RX MEDLINE=9793496; PubMed=9249990;
 RA Fukuda-Tanaka S., Hoshino A., Hataomi Y., Habu Y., Hasebe M.,
 RT Iida S.;

RT "Identification of new chalcone synthase genes for flower pigmentation
 in the Japanese and common morning glories.";
 RL Plant Cell Physiol. 38:754-758(1997).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 which can under specific conditions spontaneously isomerize into
 naringenin.

CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 OF WHICH ARE BRIGHTLY COLORED.

CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 or send an email to license@isb-sib.ch).

CC EMBL; AB001827; BAA21789.1; -.
 CC FIRM; T07799; T07799.

DR InterPro; IPR001099; N-C synthase.

DR Pfam; PF00195; Chal_still_synth; 1.

DR Pfam; PF02797; Chal_still_synth; 1.

DR ProDom; PD000453; N-C synthase; 1.

DR PROSITE; PS00441; CHALCONE_SYNTH; 1.

KW Flavonoid biosynthesis; transferase; Acyltransferase;

KW Multigene family.

FT ACT SITE 164 164 BY SIMILARITY.

SEQUENCE 389 AA; 42677 MW; 0CC81302CCBA56A CRC64;

Query Match 37.4%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 2.9;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QKAKMPWIOPTKV 18
 |||:|:|:|:|:
 DB 111 OKAIKMGOPKSKI 124

RESULT 14

CHSE_PHANI STANDARD; PRT; 389 AA.

AC 022046;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chalcone synthase E (EC 2.3.1.74) (Naringenin-chalcone synthase E)
 DE (CHS-E).
 GN CHSE.

OS Pharbitis nil (Violet) (Japanese morning glory).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Convolvulaceae; Ipomoea.
 NCBI_TaxID=35883;
 [1]

SEQUENCE FROM N.A.

RC STRAIN=CV KK/2SK-2; TISSUE=Flower buds;
 RX MEDLINE=9793496; PubMed=9249990;
 RA Fukuda-Tanaka S., Hoshino A., Hataomi Y., Habu Y., Hasebe M.,
 RT Iida S.;

RT "Identification of new chalcone synthase genes for flower pigmentation
 in the Japanese and common morning glories.";
 RL Plant Cell Physiol. 38:754-758(1997).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 which can under specific conditions spontaneously isomerize into
 naringenin.

CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 OF WHICH ARE BRIGHTLY COLORED.

CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 or send an email to license@isb-sib.ch).

CC EMBL; AB001819; BAA21788.1; -.
 CC InterPro; IPR001099; N-C synthase.

DR Pfam; PF00195; Chal_still_synth; 1.

DR Pfam; PF02797; Chal_still_synth; 1.

DR ProDom; PD000453; N-C synthase; 1.

DR PROSITE; PS00441; CHALCONE_SYNTH; 1.

KW Flavonoid biosynthesis; transferase; Acyltransferase;

KW Multigene family.

FT ACT SITE 164 164 BY SIMILARITY.

SEQUENCE 389 AA; 42685 MW; 0FC0E014B9CC0312 CRC64;

Query Match 37.4%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 2.9;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QKAKMPWIOPTKV 18
 |||:|:|:|:|:
 DB 111 OKAIKMGOPKSKI 124

RESULT 15

CHSJ_PETHY STANDARD; PRT; 389 AA.

AC P22928;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chalcone synthase J (EC 2.3.1.74) (Naringenin-chalcone synthase J).
 GN CHSJ.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:55 / Search time 15.9226 Seconds
(without alignments)
150.994 Million cell updates/sec

Title: US-09-787-070-4
139
Perfect score: 1
Sequence: 1 VYOHQAKMKPWIQPKTKVPIVRYL 25

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	139	100.0	222 1	KABOS2
2	115	82.7	223 1	KASHS2
3	115	82.7	223 2	alpha-s2-casein pr
4	115	82.7	223 2	UN0547
5	70.5	50.7	235 2	alpha-s2-casein C
6	67	48.2	182 2	alpha-s2-casein -
7	53	38.1	935 2	probable DNA damag
8	52	37.4	319 2	naringenin-chalcon
9	52	37.4	340 2	naringenin-chalcon
10	52	37.4	389 1	SYPOCJ
11	52	37.4	389 1	SYPOCJ
12	52	37.4	389 1	SYPOCJ
13	52	37.4	389 1	SYPOCJ
14	52	37.4	389 1	SYPOCJ
15	52	37.4	389 1	SYPOCJ
16	51	36.7	180 2	SYSKCD
17	50.5	36.3	471 2	adenylosuccinate 1
18	50	36.0	311 2	hypothetical prote
19	50	36.0	1877 2	hypothetical prote
20	49.5	35.6	325 2	probable phosphat
21	49.5	35.6	564 2	hypothetical prote
22	49	35.3	188 2	probable zinc meta
23	49	35.3	207 2	vascular endotheli
24	49	35.3	342 2	hypothetical prote
25	49	35.3	398 2	naringenin-chalcon
26	49	35.3	400 1	SYZMCC
27	48	34.5	392 2	stilbene synthase
28	48	34.5	392 2	stilbene synthase
29	48	34.5	392 2	stilbene synthase

30	48	34.5	392 2	SS3314	stilbene synthase
31	48	34.5	418 2	F72370	pH9-related prote
32	47	33.8	53 2	S41957	naringenin-chalcon
33	47	33.8	128 2	S33511	naringenin-chalcon
34	47	33.8	158 2	B86169	hypothetical prote
35	47	33.8	285 2	S44369	naringenin-chalcon
36	47	33.8	331 2	JO1071	naringenin-chalcon
37	47	33.8	370 2	S44368	naringenin-chalcon
38	47	33.8	333 2	S35165	naringenin-chalcon
39	47	33.8	368 1	SYSCN	naringenin-chalcon
40	47	33.8	368 1	SYSCN	naringenin-chalcon
41	47	33.8	368 1	SYSCN	naringenin-chalcon
42	47	33.8	368 1	SYSCN	naringenin-chalcon
43	47	33.8	368 2	JO2249	naringenin-chalcon
44	47	33.8	368 2	JO2259	naringenin-chalcon
45	47	33.8	368 2	S60472	naringenin-chalcon

ALIGNMENTS

RESULT 1
KABOS2
alpha-s2-casein precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 01-Sep-1981 #sequence revision 03-Feb-1994 #text_change 22-Jun-1999
C:Accession: J02008; A23087; A51438; S66626; A03107
R:Gronen, M.A.W.; Dijkhof, R.J.M.; Verstege, A.V.M.; van der Poel, J.J.
Gene 123, 187-193, 1993
A:Title: The complete sequence of the gene encoding bovine alpha2-casein.
A:Reference number: J02008; MUID:93154583; PMID:8428658
A:Accession: J02008
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-222 <GRO>
A:Cross-References: GB:M94327
R:Stewart, A.F.; Bonning, U.; Beattie, C.W.; Shah, F.; Willis, I.M.; Mackinlay, A.G.
Mol. Biol. Evol. 4, 231-241, 1987
A:Title: Complete nucleotide sequences of bovine alpha-s2- and beta-casein cDNAs: compa
A:Reference number: A93062; MUID:8818989; PMID:2833669
A:Accession: A29087
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-222 <STB>
A:Cross-References: GB:M16644; NID:G162928; EIDN:AAA30479.1; PID:G162928
R:Brignon, G.; Ribadeau Dumas, B.; Mercier, J.C.; Pellissier, J.P.; Das, B.C.
FEBS Lett. 76, 274-279, 1977
A:Title: Complete amino acid sequence of bovine alpha-s2-casein.
A:Reference number: A91438; MUID:77185633; PMID:862906
A:Contents: A allele
A:Accession: A91438
A:Molecule type: protein
A:Residues: 16-101, 'EE', 104-222 <RI>
A:Note: four fractions, previously designated s2, s3, s4, and s6, appear to have the sa
these
R:Grosclaude, F.; Joudrier, P.; Mahe, M.F.
J. Dairy Res. 46, 211-213, 1979
A:Title: A genetic and biochemical analysis of a polymorphism of bovine alpha-s2-casein
A:Reference number: A92771; MUID:79239837; PMID:46904
A:Contents: annotation; D allele
A:Note: The sequence of the D allele has a deletion of nine residues, which may be 49-5
R:Nacht, H.D.; Raída, M.; Adernann, K.; Maegert, H.U.; Forssmann, W.G.
FEBS Lett. 372, 185-188, 1995
A:Title: Casocidin-I: a casein-alpha(s2) derived peptide exhibits antibacterial activit
A:Reference number: S66626; MUID:96000204; PMID:755666
A:Accession: S66626
A:Molecule type: protein
A:Residues: 165-203 <ZUC>
A:Comment: The sequence of the A allele is shown.
C:Genetics:
A:Gene: alphas2ca
A:Map position: 6
A:Introns: 17/3; 26/3; 33/3; 47/3; 56/3; 65/3; 74/3; 82/3; 97/3; 138/3; 147/3; 156/3; 1

C:Superfamily: alpha-s2-casein
 C:Keywords: mammary gland; milk; phosphoprotein
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-222/Product: alpha-s2-casein #status experimental <MAT>
 F:23,24,25,31,71,72,73,76,144,146,158/Binding site: phosphate (Ser) (covalent) #status F

Query Match 100.0%; Score 139; DB 1; Length 222;
 Best Local Similarity 100.0%; Pred. No. 4e-13;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIQKTKVIPPVRYL 25
 DB 199 VDHOKAMKPMIQKTKVIPPVRYL 222

RESULT 2
 KASHS2
 alpha-s2-casein precursor - sheep
 C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C:Date: 31-Mar-1988 #sequence, revision 31-Mar-1988 #text_change 22-Jun-1999
 C:Accession: A25070; S17856
 R:Boisnard, M.; Petrisant, G.
 Biochimie 67, 1043-1051, 1985
 A>Title: Complete sequence of ovine alpha-s2-casein messenger RNA.
 A:Reference number: A25070; MUID:86104467; PMID:3002499
 A:Accession: A25070
 A:Molecule type: mRNA
 A:Residues: 1-223 <BO1>
 A:Cross-references: GB:X03238; NID:91238; PIDN:CAA26983.1; PID:G732894
 A>Note: 64-Aan was also found
 R:Boisnard, M.; Hue, D.; Bouniol, C.; Mercier, J.C.; Gaye, P.
 Eur. J. Biochem. 201, 633-641, 1991
 A>Title: Multiple mRNA species code for two non-allelic forms of ovine alpha-s2-casein.
 A:Reference number: S17856; MUID:92037619; PMID:1935959
 A:Accession: S17856
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-93, 'T', 95-223 <BO2>
 C:Superfamily: alpha-s2-casein
 C:Keywords: mammary gland; milk; phosphoprotein
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-223/Product: alpha-s2-casein #status predicted <KA2>
 F:23,24,25,32,55,72,73,74,77,145,147,159/Binding site: phosphate (Ser) (covalent) #status F:53,86,146,154,170,198/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 82.7%; Score 115; DB 1; Length 223;
 Best Local Similarity 84.0%; Pred. No. 1.5e-09;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIQKTKVIPPVRYL 25
 DB 199 VDHOKAMKPMIQKTKVIPPVRYL 223

RESULT 3
 JN0547
 alpha-s2-casein precursor - goat
 C:Species: Capra aegagrus hircus (domestic goat)
 C:Date: 31-Dec-1993 #sequence, revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: S33881, S33880, JN0547, S20620
 R:Bouniol, C.; Brignon, G.; Mahe, M.F.; Printz, C.
 Protein Seq. Data Anal. 5, 213-218, 1993
 A>Title: Characterization of goat allelic alpha-s2-caseins A and B: further evidence of A
 A:Reference number: S33880
 A:Accession: S33881
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-223 <BO1>
 A>Note: allele B
 A:Accession: S33880
 A:Molecule type: protein
 A:Residues: 16-78, 'E', 80-223 <BO2>
 A>Note: sequence deduced from compositional analysis of peptides

R:Bouniol, C.
 Gene 125, 235-236, 1993
 A>Title: Sequence of the goat alpha-s2-casein-encoding cDNA.
 A:Reference number: JN0547; MUID:93216130; PMID:8462880
 A:Accession: JN0547
 A:Molecule type: mRNA
 A:Residues: 1-78, 'E', 80-223 <BO3>
 A:Cross-references: EMBL:X65160; NID:9955; PIDN:CAA46278.1; PID:9956
 A>Note: allele A

Query Match 82.7%; Score 115; DB 2; Length 223;
 Best Local Similarity 84.0%; Pred. No. 1.5e-09;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIQKTKVIPPVRYL 25
 DB 199 VDHOKAMKPMIQKTKVIPPVRYL 223

RESULT 4
 146995
 alpha s2-casein C - goat
 C:Species: Capra aegagrus hircus (domestic goat)
 C:Date: 21-Feb-1997 #sequence, revision 21-Feb-1997 #text_change 13-Aug-1999
 C:Accession: 146995
 R:Bouniol, C.; Brignon, G.; Mahe, M.F.; Printz, C.
 Anim. Genet. 25, 173-177, 1994
 A>Title: Biochemical and genetic analysis of variant C of caprine alpha s2-casein (Capra
 A:Reference number: 146995; MUID:95030556; PMID:7943951
 A:Accession: 146995
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-223 <BOU>
 A:Cross-references: GB:S74171; NID:G707033; PIDN:AB32166.1; PID:G707034
 C:Superfamily: alpha-s2-casein

Query Match 82.7%; Score 115; DB 2; Length 223;
 Best Local Similarity 84.0%; Pred. No. 1.5e-09;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIQKTKVIPPVRYL 25
 DB 199 VDHOKAMKPMIQKTKVIPPVRYL 223

RESULT 5
 A48383
 alpha s2-casein - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 19-Nov-1993 #sequence, revision 18-Nov-1994 #text_change 03-May-1996
 C:Accession: A48383
 R:Alexander, L.J.; Das Gupta, N.A.; Beattie, C.W.
 Anim. Genet. 23, 365-367, 1992
 A>Title: The sequence of porcine alpha s2-casein cDNA.
 A:Reference number: A48383; MUID:92367960; PMID:1503276
 A:Accession: A48383
 A>Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-235 <ALB>
 A:Experimental source: mammary gland
 A>Note: sequence inconsistent with the nucleotide translation
 A>Note: sequence extracted from NCBI backbone (NCBIN:110884, NCBI:P:110885)
 C:Superfamily: alpha-s2-casein

Query Match 50.7%; Score 70.5; DB 2; Length 235;
 Best Local Similarity 55.6%; Pred. No. 0.0062;
 Matches 15; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

OY 1 VYOHOKAMKPMIOPTK---KVIPIYVR 24
 DB 208 VYOHOKAMKPMIHKIKTNSYIIPLNRY 234

RESULT 6

S39776
 alpha-S2-casein form b precursor - rabbit
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 13-Aug-1999
 C:Accession: S39776
 R:Davidson, S.P.; Wilde, C.J.; Tighe, P.J.; Mayer, R.J.
 Biochem. J. 296, 777-784, 1993
 A:Title: Characterization of two novel casein transcripts in rabbit mammary gland.
 A:Reference number: S39775; MUID:94107245; PMID:8280077
 A:Accession: S39776
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-182 <DNA>
 A:Cross-references: EMBL:X76909; NID:9439527; PIDN:CAA54231.1; PID:9439528
 C:Superfamily: alpha-s2-casein

Query Match 48.2%; Score 67; DB 2; Length 182;
 Best Local Similarity 43.5%; Pred. No. 0.016;
 Matches 10; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

OY 1 VYOHOKAMKPMIOPTKVIPIYVR 23
 DB 83 LYQYPTVMDPMTAEKAIPIR 105

RESULT 7

T3389
 probable DNA damage repair protein SPBC1347.01c SPBC215.16c - fission yeast (Schizosacch
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Sep-2000
 C:Accession: T3389; T33906
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
 submitted to the EMBL Data Library, February 1999
 A:Reference number: Z21851
 A:Accession: T3389
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-935 <WCO>
 A:Cross-references: EMBL:AL035548; PIDN:CA837432.1; GSPDB:GN00067; SPDB:SPBC1347.01c
 R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
 submitted to the EMBL Data Library, November 1998
 A:Reference number: Z21889
 A:Accession: T33906
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 731-935 <LYN>
 A:Cross-references: EMBL:AL033534; PIDN:CAA22130.1; GSPDB:GN00067; SPDB:SPBC215.16c
 C:Genetics:
 A:Gene: SPDB:SPBC1347.01c
 A:Map position: 2

Query Match 38.1%; Score 53; DB 2; Length 935;
 Best Local Similarity 37.0%; Pred. No. 10;
 Matches 10; Conservative 8; Mismatches 5; Indels 4; Gaps 2;

OY 2 YOHOKAMP-WIQ---PKTVIPYVR 24
 DB 119 WKHOKVPMIWDICIKOKILIPWNY 145

RESULT 8

S1816
 narigenin-chalcone synthase (EC 2.3.1.74) - garden petunia
 C:Species: Petunia x hybrida (garden petunia)
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-May-2000

C:Accession: S18136
 R:van Tunen, A.J.
 submitted to the EMBL Data Library, March 1989
 A:Reference number: S18136
 A:Accession: S18136
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-319 <TUN>
 A:Cross-references: EMBL:X14599; NID:g20537; PIDN:CAA32739.1; PID:g20538
 C:Superfamily: chalcone synthase
 C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 37.4%; Score 52; DB 2; Length 319;
 Best Local Similarity 64.3%; Pred. No. 4.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 OKAMKPMIOPTKV 18
 DB 41 OKATKMGOPKSKI 54

RESULT 9

T10957
 narigenin-chalcone synthase (EC 2.3.1.74) CHS-Fl - common morning-glory
 C:Species: Ipomoea purpurea (common morning-glory)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 05-May-2000
 C:Accession: T10957
 R:Rauscher, M.D.; Tiffin, P.L.; Miller, R.E.
 submitted to the EMBL Data Library, January 1997
 A:Description: Regulation of anthocyanin gene expression in Ipomoea purpurea.
 A:Reference number: 217229
 A:Accession: T10957
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-940 <RAU>
 A:Cross-references: EMBL:U74082; NID:g1658119; PID:g1658120
 A:Experimental source: flower bud; WMA genotype (dark flowered)
 C:Genetics:
 A:Gene: CHS-Fl
 C:Superfamily: chalcone synthase
 C:Keywords: acyltransferase; coenzyme A

Query Match 37.4%; Score 52; DB 2; Length 340;
 Best Local Similarity 64.3%; Pred. No. 5.1;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 OKAMKPMIOPTKV 18
 DB 111 OKATKMGOPKSKI 124

RESULT 10

SYJCI
 narigenin-chalcone synthase (EC 2.3.1.74) J - garden petunia
 N:Alternate names: chalcone synthase
 C:Species: Petunia x hybrida (garden petunia)
 C:Date: 30-Sep-1991 #sequence_revision 17-Mar-2000 #text_change 05-May-2000
 C:Accession: D72821; J50309
 R:Koes, R.E.; Spelt, C.B.; van den Elzen, P.J.M.; Mol, J.N.M.
 Gene 81, 245-257, 1989
 A:Title: Cloning and molecular characterization of the chalcone synthase multigene fami
 A:Reference number: J50308; MUID:90034197; PMID:2806915
 A:Accession: D72821
 A:Molecule type: DNA
 A:Residues: 1-389 <KOE>
 A:Cross-references: EMBL:X14597; NID:g20535; PIDN:CAA32737.1; PID:g20536
 A:Experimental source: strain Violet 30; leaf
 A:Accession: J50309
 A:Molecule type: DNA
 A:Residues: 1-50, 'D', '52-74', 'V', '76-228', 'I', '230-297', 'U', '299-389 <KO2>
 A:Note: the sequence is revised in Genbank entry PCHSD release 114, (PIDN:CAA32737.1)
 C:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavon

C:Genetics:

A:Gene: chsJ
 A:Map position: V
 A>Note: chsJ is expressed in various floral tissues and UV illuminated seedlings
 C:Superfamily: chalcone synthase
 C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 37.4%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 5.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTKV 18
 DB 111 OKAIKEWGPXSKI 124

RESULT 11

SYNJC
 naringenin-chalcone synthase (EC 2.3.1.74) R - garden petunia

N:Alternate names: chalcone synthase
 C:Species: Petunia x hybrida (garden petunia)
 C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000
 C:Accession: A23643

R:Koe, R.E.; Spelt, C.E.; Reif, H.J.; van den Elzen, P.J.M.; Veltkamp, R.; Mol, J.N.M.
 Nucleic Acids Res. 14, 5229-5239, 1996

A:Title: Floral tissue of Petunia hybrida (V30) expresses only one member of the chalcone synthase gene family

A:Accession: A23643
 A:Reference number: MUID:86286540; PMID:3016642

A:Molecule type: mRNA
 A:Residues: 1-389 <KOE>

A:Cross-references: GB:X04080; NID:G20541; PIDN:CAA27718.1; PID:G20542

A:Experimental source: Strain Violet 30, flowers
 A:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavonoids

A:Gene: chsR
 A:Note: expressed in floral tissue

C:Superfamily: chalcone synthase
 C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 37.4%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 5.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTKV 18
 DB 111 OKAIKEWGPXSKI 124

RESULT 12
 SYNJC
 naringenin-chalcone synthase (EC 2.3.1.74) A - garden petunia

N:Alternate names: chalcone synthase
 C:Species: Petunia x hybrida (garden petunia)
 C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000
 C:Accession: J03038

R:Koe, R.E.; Spelt, C.E.; van den Elzen, P.J.M.; Mol, J.N.M.
 Gene 81, 245-257, 1989

A:Title: Cloning and molecular characterization of the chalcone synthase multi-gene family

A:Accession: J03038
 A:Reference number: MUID:90034157; PMID:2806915

A:Molecule type: DNA
 A:Residues: 1-389 <KOE>

A:Cross-references: GB:X14591; NID:G20524; PIDN:CAA32731.1; PID:G20525

A:Experimental source: Strain Violet 30, leaf
 A:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavonoids

A:Gene: chsA
 A:Map position: V
 A:Note: chsA is the major expressed member of the gene family in various floral tissues

C:Superfamily: chalcone synthase
 C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 37.4%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 5.8;

Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTKV 18
 DB 111 OKAIKEWGPXSKI 124

RESULT 13
 SYNJC
 naringenin-chalcone synthase (EC 2.3.1.74) 2 - potato

N:Alternate names: chalcone synthase; CHS
 C:Species: Solanum tuberosum (potato)
 C:Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 05-May-2000
 C:Accession: J05136; PC4239

R:Jeon, J.H.; Kim, H.S.; Choi, K.H.; Young, Y.H.; Young, H.; Byun, S.M.
 Biosci. Biotechnol. Biochem. 60, 1907-1910, 1996

A:Title: Cloning and characterization of one member of the chalcone synthase gene family

A:Accession: J05136
 A:Reference number: MUID:97141614; PMID:8987872

A:Molecule type: mRNA
 A:Residues: 1-389 <JEO1>

A:Cross-references: GB:U47738; NID:G1470059; PIDN:AAH05239.1; PID:G1470060

A:Accession: PC4239
 A:Status: preliminary

A:Molecule type: Protein
 A:Residues: 158-165/367-373 <JEO2>

A:Comment: This enzyme is important in the biosynthesis of all classes of flavonoids in C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A

Query Match 37.4%; Score 52; DB 2; Length 389;
 Best Local Similarity 64.3%; Pred. No. 5.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTKV 18
 DB 111 OKAIKEWGPXSKI 124

RESULT 14
 SYNJC
 naringenin-chalcone synthase (EC 2.3.1.74) - common morning-glory

N:Alternate names: chalcone synthase
 C:Species: Ipomoea purpurea (common morning-glory)
 C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
 C:Accession: T07799

R:Fukada-Tanaka, S.; Hoshino, A.; Hisatomi, Y.; Habu, Y.; Hasebe, M.; Iida, S.
 Plant Cell Physiol. 38, 754-758, 1997

A:Title: Identification of new chalcone synthase genes for flower pigmentation in the J

A:Reference number: Z16440; MUID:97393496; PMID:9249990

A:Accession: T07799
 A:Status: preliminary; translated from GB/EMBL/DBU

A:Molecule type: mRNA
 A:Residues: 1-389 <FUK>

A:Cross-references: EMBL:AB01827; NID:G2329836; PIDN:BAA21789.1; PID:G2329837

A:Gene: CHD-E
 C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 37.4%; Score 52; DB 2; Length 389;
 Best Local Similarity 64.3%; Pred. No. 5.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTKV 18
 DB 111 OKAIKEWGPXSKI 124

RESULT 15
 SYNJC
 naringenin-chalcone synthase (EC 2.3.1.74) - garden snapdragon

N:Alternate names: chalcone synthase
 C:Species: Antirrhinum majus (garden snapdragon)
 C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000
 C:Accession: S07312; A33217
 R:Sommer, H.; Saedler, H.
 M:1. Gen. Genet. 202, 429-434, 1986
 A>Title: Structure of the chalcone synthase gene of Antirrhinum majus.
 A:Reference number: S07312
 A:Accession: S07312
 A:Molecule type: DNA
 A:Residues: 1-390 <SCM>
 A:Cross-references: EMBL:X03710; NID:G16015; PIDN:CAA27338.1; PID:G16016
 A:Accession: A33217
 A>Status: translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-390 <SCM2>
 C:Genetics:
 A:Gene: chs
 A:Introns: 60/1; 162/3
 C:Superfamily: chalcone synthase
 C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 37.4%; Score 52; DB 1; Length 390;
 Best local similarity 64.3%; Pred. No. 5.9;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKMKPWLOPKTY 18
 DB 111 OKAIKEWQPKSKI 124

Search completed: July 30, 2003, 16:31:15
 Job time : 15.9226 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using SW model

Run on: July 30, 2003, 16:09:50 ; Search time 49.256 Seconds
(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-4
Perfect score: 139
Sequence: 1 VYGHOKAMKPWIQPKTKVIVYRYL 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database :

A_Geneseq_13Jun03.*

1:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	139	100.0	25	13	AAK20055	Casain peptide. S
2	139	100.0	25	21	AAI88340	C-terminal peptide
3	139	100.0	31	18	AAW32219	Alpha-52 casein pr
4	139	100.0	31	23	AAE17466	Bovine alpha-s2 ca
5	139	100.0	222	18	AAW32220	Bovine alpha-52 ca
6	139	100.0	232	23	AAE17468	Bovine alpha-s2 ca
7	135	97.1	24	15	AAE60481	peptide used in cr
8	135	97.1	24	21	AAE88339	peptide used in cr
9	128.5	92.4	26	20	AAW93885	Bifidobacterium B

10	119	85.6	21	21	AAE8338	Internal peptide f
11	115	82.7	223	23	AAE17469	Alpha-S2 casein pr
12	115	82.7	223	23	AAE17470	Coat alpha-S2 case
13	115	82.7	223	23	AAE17471	Capra hircus alpha
14	115	82.7	223	23	AAE17473	Sheep alpha-S2 cas
15	115	82.7	223	23	AAE17475	Alpha-S2 casein pr
16	105	75.5	19	18	AAW32218	Alpha-S2 casein pr
17	105	75.5	19	21	AAE17465	Internal alpha-S2 ca
18	92	66.2	16	21	AAE188337	Internal peptide f
19	70.5	50.7	234	23	AAE17474	Pig alpha-S2 casei
20	69	49.6	13	23	AAE17464	Bovine alpha-S2 ca
21	67	48.2	178	23	AAE17472	Rabbit alpha-S2 ca
22	64	46.0	112	23	AAE17463	Bovine alpha-S2 ca
23	57	41.0	11	18	AAW32217	Alpha-S2 casein pr
24	57	41.0	11	23	AAE17462	Bovine alpha-S2 ca
25	53	38.1	9	17	AAW04282	Milk derived anti-
26	52	37.4	10	18	AAW32216	Alpha-S2 casein pr
27	52	37.4	10	23	AAE17461	Bovine alpha-S2 ca
28	52	37.4	360	22	AAE60169	Patina caseinase s
29	52	37.4	389	17	AAE60165	Chalcone synthase
30	52	37.4	389	22	AAE74019	Chalcone synthase
31	52	37.4	389	22	AAE73991	Cyclamen chalcone
32	51	36.7	394	24	AAE60374	Petunia chalcone s
33	51	36.7	399	24	AAE60375	Humulus lupulus ch
34	50.5	36.3	471	21	AAE18201	Humulus lupulus ch
35	50	36.0	291	17	AAE69092	Plasmodium falci
36	49	35.3	8	12	AAW04278	Milk derived anti-
37	49	35.3	133	17	AAW04828	Milk cGV protein s
38	49	35.3	133	20	AAW04922	Milk cGV protein s
39	49	35.3	133	23	AAU83406	Vascular endotheli
40	49	35.3	160	20	AAW66208	Vascular endotheli
41	49	35.3	167	20	AAW66224	Human VEGF-B full
42	49	35.3	188	17	AAW04826	Heart vascular end
43	49	35.3	188	17	AAW00864	Murine VEGF167. Mu
44	49	35.3	188	20	AAW66201	Human vascular end
45	49	35.3	188	20	AAW08490	Murine vascular en

ALIGNMENTS

XX	RESULT 1
AC	ARR20055
ID	ARR20055 strand; peptide; 25 AA.
XX	ARR20055;
XX	25-MAR-2003 (updated)
DT	26-MAR-1992 (first entry)
DT	
XX	Casein peptide.
DE	
XX	Casein; platelet; aggregation; thrombosis; collagen.
KM	
XX	Synthetic.
OS	
XX	JF03255095-A.
PN	
XX	13-NOV-1991.
PD	
XX	02-MAR-1990; 90JP-0052554.
PF	
XX	02-MAR-1990; 90JP-0052554.
PR	
XX	(KANE) KANEBO LTD.
XX	
XX	WPI; 1992-002669/01.
DR	
XX	Casein peptide(s) for treating thrombosis - as inhibitors of
PT	platelet aggregation caused by adenosine-5'-di-phosphate and
PT	collagen and as biochemical reagents
XX	
XX	Claim 1; Page 1; 8pp; Japanese.
PS	

XX Leu25 may be omitted. The peptide and its salts have inhibiting
 CC activity against platelet aggregation caused by adenosine-5'-
 CC diphosphate and by collagen. They are useful for prevention and
 CC treatment of thrombosis and are also useful as a biochemical reagent.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX

SQ Sequence 25 AA;

Query Match 100.0%; Score 139; DB 13; Length 25;
 Best Local Similarity 100.0%; Pred. No. 2e-13; Indels 0; Gaps 0;
 Matches 25; Conservative 0; Mismatches 0;

QY 1 VYQHOKAMKPMIQPKTKVIPYVRYL 25
 |||||
 DB 1 VYQHOKAMKPMIQPKTKVIPYVRYL 25

RESULT 2

AAI8340 standard; peptide; 25 AA.

XX AAY88340;

DT 14-JUL-2000 (first entry)

XX C-terminal peptide fragment of bovine alpha-S2 casein.

XX Alpha-S2 casein; peptide production; biological fluid; milk; whey; blood;
 KW antibacterial peptide; lactoferrin; antiviral; antitumour activity.

XX Bos sp.

PN MO200015655-A1.

XX 23-MAR-2000.

PF 15-SEP-1999; 99WO-EP07002.

PR 15-SEP-1998; 98EP-0203107.

PR 08-JUN-1999; 99EP-0201815.

PA (NIZO-) NIZO FOOD RES.

PI Visser S, Recio I;

DR WPI; 2000-271377/23.

PT Novel process for producing peptides with e.g. antimicrobial activity
 PT from biological fluids such as milk, whey or blood comprises contacting
 PT fluid with chromatographic medium to adsorb peptide domain of interest

PS Claim 14; Page 22; 41pp; English.

CC This sequence represents a C-terminal fragment of bovine alpha-S2 casein
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptide obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.

SQ Sequence 25 AA;

Query Match 100.0%; Score 139; DB 21; Length 25;
 Best Local Similarity 100.0%; Pred. No. 2e-13; Indels 0; Gaps 0;
 Matches 25; Conservative 0; Mismatches 0;

QY 1 VYQHOKAMKPMIQPKTKVIPYVRYL 25
 |||||
 DB 1 VYQHOKAMKPMIQPKTKVIPYVRYL 25

RESULT 3

AAI32219 standard; peptide; 31 AA.

XX AAM32219;

DT 03-FEB-1998 (first entry)

XX Alpha-S2 casein precursor C-terminal peptide fragment #5.

XX Alpha-S2 casein precursor; growth promoting; mitogenic assay;
 KW platelet-derived growth factor; insulin-like derived growth factor;

XX Synthetic.

OS Bos taurus.

PN MO9716460-A1.

PD 09-MAY-1997.

PF 31-OCT-1996; 96WO-GB02658.

PR 31-OCT-1995; 95GB-0022302.

PA (UNIV-) UNIV LIVERPOOL.

PI Liu Q, Smith JA, Wilkinson MC;

DR WPI; 1997-272048/24.

PT Manufacture of medicament or foodstuff for promoting growth - using
 PT peptide(s) with a sequence identical to the C-terminal end of an
 PT alpha-S2 casein precursor

PS Claim 12; Page 21; 33pp; English.

CC The present sequence, which is substantially identical to the C-terminal
 CC end of an alpha-S2 casein precursor, was found after storage in PBS to
 CC exhibit growth promoting activity for rat mammary fibroblast cell line
 CC (Rama 27), which is not significantly stimulated by platelet-derived
 CC growth factor or insulin-like growth factor. The activity of the peptide
 CC increased when maintained at alkaline pH. By way of contrast, alpha-S2
 CC casein was inactive in a mitogenic assay. This peptide may be used in the
 CC manufacture of a medicament or foodstuff for promoting growth in
 CC humans or animals.

SQ Sequence 31 AA;

Query Match 100.0%; Score 139; DB 18; Length 31;
 Best Local Similarity 100.0%; Pred. No. 2.5e-13; Indels 0; Gaps 0;
 Matches 25; Conservative 0; Mismatches 0;

QY 1 VYQHOKAMKPMIQPKTKVIPYVRYL 25
 |||||
 DB 7 VYQHOKAMKPMIQPKTKVIPYVRYL 31

RESULT 4

AAI17466 standard; peptide; 31 AA.

XX AAI17466;

XX

DT 22-APR-2002 (first entry)
 XX
 DE Bovine alpha-S2 casein precursor protein C-terminal fragment #7.
 XX
 KM Bovine; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KM chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Bos sp.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 PI
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 PS
 PS Claim 9; Page 22; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptide stimulates the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is a peptide
 CC derived from the C-terminal of bovine alpha-S2 casein precursor protein.
 CC
 SQ Sequence 31 AA;
 Query Match 100.0%; Score 139; DB 23; Length 31;
 Best Local Similarity 100.0%; Pred. No. 2.5e-13;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKPMIQPKTKVIPYRYL 25
 DB 7 VYQHOKAMKPMIQPKTKVIPYRYL 31
 RESULT 5
 AAM32220
 ID AAM32220 standard; protein; 222 AA.
 XX
 AC AAM32220;
 XX
 DT 03-FEB-1998 (first entry)
 XX
 DE Bovine alpha-S2 casein precursor.
 XX
 KM Bovine alpha-S2 casein precursor; growth promoting; mitogenic assay;
 KM platelet-derived growth factor; insulin-like derived growth factor;
 XX
 OS Bos taurus.
 XX
 PN WO9716460-A1.
 XX
 PD 09-MAY-1997.
 XX

PF 31-OCT-1996; 96WO-GB02658.
 XX
 PR 31-OCT-1995; 95GB-0022302.
 XX
 PA (UWLI-) UNIV LIVERPOOL.
 XX
 PI Liu Q, Smith JA, Wilkinson MC;
 XX
 DR WPI; 1997-272048/24.
 XX
 PT Manufacture of medicament or foodstuff for promoting growth - using
 PT peptide(s) with a sequence identical to the C-terminal end of an
 PT alpha-S2 casein precursor
 XX
 PS Disclosure; Page 3; 33pp; English.
 XX
 CC The present sequence represents bovine alpha-S2 casein precursor.
 CC Peptides having an amino acid sequence which is substantially identical
 CC to the C-terminal end of an alpha-S2 casein precursor, are used for
 CC manufacture of a medicament or foodstuff for promoting growth in humans
 CC or animals.
 CC
 SQ Sequence 222 AA;
 Query Match 100.0%; Score 139; DB 18; Length 222;
 Best Local Similarity 100.0%; Pred. No. 2.3e-12;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKPMIQPKTKVIPYRYL 25
 DB 198 VYQHOKAMKPMIQPKTKVIPYRYL 222
 RESULT 6
 AAE17468
 ID AAE17468 standard; Protein; 222 AA.
 XX
 AC AAE17468;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Bovine alpha-S2 casein precursor protein.
 XX
 KM Bovine; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KM chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Bos sp.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 PI
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 PS
 PS Claim 8; Page 6; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and

CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is bovine
 CC alpha-S2 casein precursor protein.

CC Sequence 222 AA;

Query Match 100.0%; Score 139; DB 23; Length 222;
 Best Local Similarity 100.0%; Pred. No. 2.3e-12;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPKTKVPIPVRYL 25
 |||||
 DB 198 VYOHOKAMKPMIOPKTKVPIPVRYL 222

RESULT 7

ID AAR60481 standard; peptide; 24 AA.

XX AAR60481;

DT 19-APR-1995 (first entry)

XX Peptide used in tranquilliser.

XX Peptide; tranquilliser; food; foodstuff; mental stabilisation.

OS Synthetic.

XX JP06211689-A.

PD 02-AUG-1994.

PF 19-JAN-1993; 93JP-0024811.

PR 19-JAN-1993; 93JP-0024811.

XX (KANE) KANEBO LTD.

DR WPI; 1994-283276/35.

PT Synthetic peptide(s) used in tranquilliser - also used in the
 production of a food for mental stabilisation.

XX Claim 3; Page 1; 5pp; Japanese.

CC The peptide may be used as a component of a tranquilliser which may
 CC be administered orally and is low in toxicity. The peptide may also
 CC be used as a component of a foodstuff which may be used for mental
 CC stabilisation. See also AAR60480.

XX Sequence 24 AA;

Query Match 97.1%; Score 135; DB 15; Length 24;
 Best Local Similarity 100.0%; Pred. No. 7.3e-13;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPKTKVPIPVRY 24
 |||||
 DB 1 VYOHOKAMKPMIOPKTKVPIPVRY 24

RESULT 8

ID AAY88339 standard; peptide; 24 AA.

XX AAY88339;

XX

DT 14-JUL-2000 (first entry)

XX Internal peptide fragment of bovine alpha-S2 casein #3.

XX Alpha-S2 casein; peptide production; biological fluid; milk; whey; blood;

XX antibacterial peptide; lactoferrin; antiviral; antitumour activity.

OS Bos sp.

XX WO200015655-A1.

PD 23-MAR-2000.

PF 15-SEP-1999; 99WO-EP07002.

PR 15-SEP-1998; 98EP-0203107.

PR 08-JUN-1999; 99EP-0201815.

XX (NIZO-) NIZO FOOD RES.

PI Visser S, Recio I;

DR WPI; 2000-271377/23.

PT Novel process for producing peptides with e.g. antimicrobial activity
 from biological fluids such as milk, whey or blood comprising contacting
 PT fluid with chromatographic medium to adsorb peptide domain of interest

XX Claim 14; Page 22; 41pp; English.

CC This sequence represents an internal fragment of bovine alpha-S2 casein
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.

XX Sequence 24 AA;

Query Match 97.1%; Score 135; DB 21; Length 24;
 Best Local Similarity 100.0%; Pred. No. 7.3e-13;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPKTKVPIPVRY 24
 |||||
 DB 1 VYOHOKAMKPMIOPKTKVPIPVRY 24

RESULT 9

ID AAW93885 standard; peptide; 26 AA.

XX AAW93885;

DT 25-JUN-1999 (first entry)

DE Bifidobacterium bifidus stimulating peptide 21.

XX Bifidogenic peptide; protease; treatment; microbe-related disease;

XX bacteria; fungi; yeast; protozoa; virus; mycoplasma; filaria; plasmodia;

XX infection; inflammation; microbial induced tumour; degenerative disorder;

XX diarrhoea; colic; oral microflora; intestinal microflora; caries;

KW vaginal microflora.
 OS Bifidobacterium bifidus.
 PN WO9914231-A2.
 XX
 PD 25-MAR-1999.
 XX
 PF 16-SEP-1998; 98WO-EP05899.
 XX
 PR 11-FEB-1998; 98DE-1005385.
 PR 16-SEP-1997; 97DE-1040604.
 XX
 PA (FORS/) FORSSMANN W.
 PI Forssmann W, Liepke C, Zucht H;
 XX WPI, 1999-244022/20.
 DR
 XX
 PT Milk-derived peptides that stimulate Bifidobacterium bifidus
 PS Claim 2; Page 3; 25pp; German.
 XX
 CC This invention describes milk-derived bifidogenic peptides and their
 CC active derivatives or fragments, and combinations of them produced by
 CC chemical coupling. Such are produced from bovine or human milk by
 CC treatment for 2 hr with proteases, then centrifuging to remove fat and
 CC acidifying to pH 2 to precipitate proteins. The solution phase is then
 CC subjected to reverse-phase high-performance liquid chromatography (HPLC)
 CC and cation-exchange HPLC, the fractions adjusted to salt content below
 CC 25 mM (by dialysis or reverse-phase HPLC) and tested for activity by
 CC growing Bifidobacterium bifidus and Escherichia coli in presence of the
 CC fractions. Those fractions for which (Bw-B0)-(Bw-B0) is at least 0.15
 CC are selected where Bw = germ count after 16 hr culture of B. bifidus in
 CC 50% Elliker broth containing peptide at 0.2 mg/ml, B0 = germ count under
 CC similar conditions in a peptide-free control, Bw = germ count after 16
 CC hr culture of E. coli in 3 g/l tryptic broth containing peptide at
 CC 0.2 mg/ml, B0 = germ count under similar conditions in a peptide-free
 CC control. The peptides AAW93865-W93888 are used to treat microbe-related
 CC diseases caused by bacteria, fungi, yeast, protozoa, viruses,
 CC mycoplasmas, filaria and plasmodia, e.g. infections, inflammation,
 CC microbially induced tumours or degenerative disorders, diarrhoea, colic,
 CC abnormalities in oral, intestinal or vaginal microflora, or carrier.
 CC
 SQ Sequence .26 AA;
 XX
 XX
 Query Match 92.4%; Score 128.5; DB 20; Length 26;
 Best Local Similarity 96.2%; Pred. No. 7.2e-12;
 Matches 25; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 VYOHOKAM-KRWIOPKTKVTPYRYL 25
 DB 1 VYOHOKAMKRWIOPKTKVTPYRYL 26
 XX
 RESULT 10
 AAY88338
 ID AAY88338 standard; peptide; 21 AA.
 AC AAY88338;
 XX
 DT 14-JUN-2000 (first entry)
 XX
 DB Internal peptide fragment of bovine alpha-S2 casein #2.
 XX
 KW Alpha-S2 casein; peptide production; biological fluid; milk; whey; blood;
 KW antibacterial peptide; lactoferrin; antiviral; antitumour activity.
 XX
 OS Bos sp.
 XX WO200015655-A1.
 PN
 XX 23-MAR-2000.
 PD

XX
 PF 15-SEP-1999; 99WO-EP07002.
 XX
 PR 15-SEP-1998; 98EP-0203107.
 PR 08-JUN-1999; 99EP-0201815.
 XX
 PA (NIZO-) NIZO FOOD RES.
 XX
 PI Visser S, Recio I;
 XX WPI, 2000-271377/23.
 DR
 XX
 PT Novel process for producing peptides with e.g. antimicrobial activity
 PT from biological fluids such as milk, whey or blood comprises contacting
 PT fluid with chromatographic medium to adsorb peptide domain of interest
 PT
 PS Claim 14; Page 22; 41pp; English.
 XX
 CC This sequence represents an internal fragment of bovine alpha-S2 casein
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin. Using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.
 CC
 SQ Sequence 21 AA;
 XX
 XX
 Query Match 85.6%; Score 119; DB 21; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYOHOKAMKRWIOPKTKVTPY 21
 DB 1 VYOHOKAMKRWIOPKTKVTPY 21
 XX
 RESULT 11
 AAE17469
 ID AAE17469 standard; Protein; 223 AA.
 AC AAE17469;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DB Alpha-S2 casein precursor (alpha-S2-CN) protein #1.
 XX
 KW Alpha-S2 casein; alpha-S2-CN; dermatological; antiinflammatory; cosmetic;
 KW fibroblast; collagen; keratinocyte; skin regeneration; medicament; aging;
 KW toothpaste; chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Unidentified.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX

PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Disclosure; Page 8; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is alpha-S2
 CC casein precursor (alpha-S2-CN) protein.
 CC
 XX Sequence 223 AA;
 SQ
 Query Match 82.7%; Score 115; DB 23; Length 223;
 Best Local Similarity 84.0%; Pred. No. 7.5e-09;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 1 VYQHOKAMKPMIOPKTKVPIPYRYL 25
 DB 199 VDOHQAMKPMIOPKTKVPIPYRYL 223
 RESULT 12
 ID AAE17470
 AC AAE17470 standard; Protein; 223 AA.
 XX
 AC AAE17470;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Goat alpha-S2 casein E precursor protein.
 XX
 KW Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KW chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Capra hircus.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Claim 8; Page 8; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC derivative. The peptide contains an amino acid sequence from alpha-S2

CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat
 CC alpha-S2 casein E precursor protein.
 CC
 XX Sequence 223 AA;
 SQ
 Query Match 82.7%; Score 115; DB 23; Length 223;
 Best Local Similarity 84.0%; Pred. No. 7.5e-09;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 1 VYQHOKAMKPMIOPKTKVPIPYRYL 25
 DB 199 VDOHQAMKPMIOPKTKVPIPYRYL 223
 RESULT 13
 ID AAE17471
 AC AAE17471 standard; Protein; 223 AA.
 XX
 AC AAE17471;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Capra hircus alpha-S2 casein C precursor protein.
 XX
 KW Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KW chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Capra hircus.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Claim 8; Page 8; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat
 CC alpha-S2 casein C precursor protein.
 CC
 XX Sequence 223 AA;
 SQ
 Query Match 82.7%; Score 115; DB 23; Length 223;

Best Local Similarity 84.0%; Pred. No. 7.5e-09;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKAMKRWIOPTKVIPYRYL 25
DB 199 VDOHQKAMKRWIOPTKVIPYRYL 223

RESULT 14
AAE17473

ID AAE17473 standard; Protein; 223 AA.

AC AAE17473;

DT 22-APR-2002 (first entry)

DE Sheep alpha-S2 casein precursor protein.

KW Sheep; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
chewing gum; cosmetic; wrinkling; periodontal disease.

OS Ovis sp.

PN WO200202133-A2.

PD 10-JAN-2002.

PF 13-JUN-2001; 2001WO-GB02601.

PR 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD.

PI Smith JA;

DR WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in
alpha-S2 casein precursor in the manufacture of a medicament for
alleviating or preventing periodontal disease and an effect of aging in
skin

PS Claim 8; Page 9; 27pp; English.

CC The invention relates to a composition comprising a peptide or its
derivative. The peptide contains an amino acid sequence from alpha-S2
casein precursor. The peptides stimulate the growth of fibroblasts, and
thus the synthesis and secretion of collagen. The peptides also stimulate
the growth of keratinocytes, which aid in the formation and regeneration
of skin surface. The peptide is useful in the manufacture of a medicament
in the form of a toothpaste or a chewing gum, for alleviating or
preventing periodontal disease and a medicament in the form of a cosmetic
composition for alleviating or preventing an effect of aging,
particularly wrinkling of the skin. The present sequence is sheep
alpha-S2 casein precursor protein.

SQ Sequence 223 AA;

Query Match 82.7%; Score 115; DB 23; Length 223;

Best Local Similarity 84.0%; Pred. No. 7.5e-09;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKAMKRWIOPTKVIPYRYL 25
DB 199 VDOHQKAMKRWIOPTKVIPYRYL 223

RESULT 15

ID AAE17475 standard; Protein; 223 AA.

AC AAE17475;

DT 22-APR-2002 (first entry)

DE Alpha-S2 casein precursor (alpha-S2-CN) protein #2.

KW Alpha-S2 casein; alpha-S2-CN; dermatological; antiinflammatory; cosmetic;
fibroblast; collagen; keratinocyte; skin regeneration; medicament; aging;
toothpaste; chewing gum; cosmetic; wrinkling; periodontal disease.

OS Unidentified.

PN WO200202133-A2.

PD 10-JAN-2002.

PF 13-JUN-2001; 2001WO-GB02601.

PR 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD.

PI Smith JA;

DR WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in
alpha-S2 casein precursor in the manufacture of a medicament for
alleviating or preventing periodontal disease and an effect of aging in
skin

PS Disclosure; Page 9; 27pp; English.

CC The invention relates to a composition comprising a peptide or its
derivative. The peptide contains an amino acid sequence from alpha-S2
casein precursor. The peptides stimulate the growth of fibroblasts, and
thus the synthesis and secretion of collagen. The peptides also stimulate
the growth of keratinocytes, which aid in the formation and regeneration
of skin surface. The peptide is useful in the manufacture of a medicament
in the form of a toothpaste or a chewing gum, for alleviating or
preventing periodontal disease and a medicament in the form of a cosmetic
composition for alleviating or preventing an effect of aging,
particularly wrinkling of the skin. The present sequence is alpha-S2
casein precursor (alpha-S2-CN) protein.

SQ Sequence 223 AA;

Query Match 82.7%; Score 115; DB 23; Length 223;

Best Local Similarity 84.0%; Pred. No. 7.5e-09;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKAMKRWIOPTKVIPYRYL 25
DB 199 VDOHQKAMKRWIOPTKVIPYRYL 223

Search completed: July 30, 2003, 16:23:27
Job time : 49.256 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 / Search time 49.1964 Seconds
(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-5

Perfect score: 164
Sequence: 1 PEMSKCYQWQRMKXGAPSTICIRRTSA 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	61.0	234	4	Q8IX02
2	99	60.4	711	4	Q8ICD2
3	97	59.1	711	4	Q8ICD6
4	97	59.1	711	4	Q8ICD6
5	97	59.1	711	4	Q8ICD6
6	97	59.1	711	4	Q8ICD6
7	97	59.1	711	4	Q8ICD6
8	97	59.1	711	4	Q8ICD6
9	97	59.1	711	4	Q8ICD6
10	97	59.1	711	4	Q8ICD6
11	97	59.1	711	4	Q8ICD6
12	97	59.1	711	4	Q8ICD6
13	97	59.1	711	4	Q8ICD6
14	97	59.1	711	4	Q8ICD6
15	97	59.1	711	4	Q8ICD6
16	97	59.1	711	4	Q8ICD6

17	53	32.3	1449	12	Q65974	Q65974	casava com
18	52	31.7	189	11	Q6R011	Q6R011	mus musculus
19	51.5	31.4	1795	4	Q14674	Q14674	homo sapien
20	51.5	31.4	4437	2	Q8GAP3	Q8GAP3	streptomyces
21	51	31.1	489	16	P96223	P96223	mycobacteri
22	50	30.5	54	6	Q9TQV7	Q9TQV7	equus caball
23	50	30.5	154	6	Q8PF84	Q8PF84	xanthomonas
24	50	30.5	511	11	Q8BVH0	Q8BVH0	mus musculus
25	50	30.5	552	5	Q9VW14	Q9VW14	dirosophila
26	50	30.5	643	11	Q8BV16	Q8BV16	mus musculus
27	50	30.5	1207	4	Q8TB59	Q8TB59	homo sapien
28	49.5	30.2	1536	4	Q9C014	Q9C014	homo sapien
29	49	29.9	130	17	Q8TVE4	Q8TVE4	methanopyru
30	49	29.9	155	12	Q91W22	Q91W22	hydrangea r
31	49	29.9	258	5	Q9VAF6	Q9VAF6	listeria in
32	49	29.9	405	16	Q92F81	Q92F81	listeria in
33	49	29.9	408	16	Q8YAE4	Q8YAE4	listeria mo
34	49	29.9	507	16	Q92F88	Q92F88	caulobacter
35	49	29.9	558	2	Q936J3	Q936J3	rhodospirill
36	49	29.9	1721	5	Q93547	Q93547	caenorhabdi
37	48.5	29.6	71	16	Q8XJ54	Q8XJ54	yeastina pe
38	48.5	29.6	132	16	Q8C1Q9	Q8C1Q9	streptomyces
39	48.5	29.6	518	16	Q911H4	Q911H4	streptomyces
40	48	29.3	271	16	Q98J98	Q98J98	rhizobium l
41	48	29.3	352	7	Q9XS13	Q9XS13	autonomocara
42	48	29.3	352	16	Q8XQ97	Q8XQ97	raletonia s
43	48	29.3	353	10	Q92TF0	Q92TF0	oryza sativ
44	48	29.3	402	10	Q92FR8	Q92FR8	oryza sativ
45	48	29.3	424	16	Q98FV3	Q98FV3	rhizobium l

ALIGNMENTS

RESULT 1	ID	Q8IX02	PRELIMINARY;	PRT;	234 AA.
AC	Q8IX02;				
DT	01-MAR-2003 (TREMBLrel. 23, Created)				
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)				
DE	Lactoferrin (Fragment).				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Lin Z., Li N.;				
RT	"Human lactoferrin gene, 5' regulation region and exons 1-6.";				
RL	Submitted (Mar-2002) to the EMBL/GenBank/DBO databases.				
DR	EMBL; AF508798; AAC14586.1; -				
FT	NON_TER				
FT	234				
FT	234				
SQ	SEQUENCE	234 AA;	25671 MW;	2A08DAFD6B8780C7	CRC64;
Query Match	Best Local Similarity	61.0%;	Score 100;	DB 4;	Length 234;
Matches	16;	Conservative	6;	Mismatches	6;
Indels	0;	Gaps	0;		
Qy	1	PEMSKYQWQRMKXGAPSTICIRRTS 28			
Db	34	PEATKCFQWQRMKXGAPVSCIKRDS 61			
RESULT 2	ID	Q8ICD2	PRELIMINARY;	PRT;	711 AA.
AC	Q8ICD2;				
DT	01-JUN-2002 (TREMBLrel. 21, Created)				
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)				
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)				
DE	Lactoferrin.				
OS	Homo sapiens (Human).				


```
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_TaxID=10090;
RP [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Bone;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA "The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002);
DR EMBL: AK036481; BAC29450.1;
SQ SEQUENCE 707 AA; 77837 MW; E1B32F5FD8748A0F CRC64;

Query Match 51.2%; Score 84; DB 11; Length 707;
Best Local Similarity 44.4%; Pred. No. 0.00044;
Matches 12; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

OY 2 EMSKCYOMORMRKLGAPSTICRTS 28
Db 33 BEEKCLMOMEMKRVGGPPLSCVKKSS 59

RESULT 7
OBMMNB PRELIMINARY; PRT; 704 AA.
ID OBMMNB.
AC OBMMNB.
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Lactoferrin (Fragment).
LN LTF.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Mammary gland;
RA Wang S.-R., Lin T.-Y., Wang C.-N.;
RT "Isolation and expression of porcine milk lactoferrin.";
RT Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TRANSFERRIN ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION. USUALLY BICARBONATE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC EMBL: U77887; AAL40161.1; -.
DR InterPro: IPR001156; Transferrin.
DR Pfam: PF00405; transferrin; 2.
DR PRINTS: PR00422; TRANSFERRIN.
DR SMART: SM00094; TR_FER_2.
DR PROSITE: PS00205; TRANSFERRIN_1; 2.
DR PROSITE: PS00206; TRANSFERRIN_2; 2.
DR PROSITE: PS00207; TRANSFERRIN_3; 1.
KW Glycoprotein; Iron transport; Metal-binding; Transport.
FT NON TER 1
FT SEQUENCE 704 AA; 77681 MW; 64EE769F7503CC32 CRC64;

Query Match 42.1%; Score 69; DB 6; Length 704;
Best Local Similarity 51.9%; Pred. No. 0.077;
Matches 14; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

OY 2 EMSKCYOMORMRKLGAPSTICRTS 28
Db 34 EYKCRQWQSGKIRTN--PIFCIRRAS 59

RESULT 8
OQUCYS PRELIMINARY; PRT; 38 AA.
ID OQUCYS.
AC OQUCYS.
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
```

```
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Lactoferrin homolog (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96081633; PubMed=8551695;
RA Sato I.;
RT "Characterization of the 84-kDa protein with ABH activity in human
RT seminal plasma.";
RL Upn. J. Legal Med. 49:281-293 (1995).
DR HSSP: O77811; 1B1X.
DR InterPro: IPR001156; Transferrin.
DR Pfam: PF00405; transferrin; 1.
SQ SEQUENCE 38 AA; 4459 MW; 0402F490B5EBDB8 CRC64;

Query Match 41.5%; Score 68; DB 4; Length 38;
Best Local Similarity 50.0%; Pred. No. 0.006;
Matches 12; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 1 PEMSKCYOMORMRKLGAPSTICI 24
Db 15 PEADKXKQWQMRNKRKVGXPVXSI 38

RESULT 9
Q9DBD0 PRELIMINARY; PRT; 700 AA.
ID Q9DBD0.
AC Q9DBD0.
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE 130001702Rik protein.
DB 130001702Rik protein.
GN 130001702Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukumitsu Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Hono H., Kaenaka T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Pleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuell P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schmitt L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gusticich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima U., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Wilming L.,
RA Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK005035; BAB23762.1; -.
DR HSSP: P19134; 1TFD.
DR MGD: MGI:1919025; 130001702Rik.
DR InterPro: IPR001156; Transferrin.
DR Pfam: PF00405; transferrin; 2.
DR PRINTS: PR00422; TRANSFERRIN.
DR SMART: SM00094; TR_FER_2.
DR PROSITE: PS00205; TRANSFERRIN_1; 1.
DR PROSITE: PS00207; TRANSFERRIN_3; 1.
```


ID 08KPR9 PRELIMINARY; PRT; 821 AA.
 AC 08KPR9;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Hypothetical protein.
 GN SEB0025.
 OS *Synechococcus* sp. (strain PCC 7942) (*Anacystis nidulans* R2).
 OC Bacteria; Cyanobacteria; Chroococcales; *Synechococcus*.
 OX NCBI_TaxID=1140;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PCC 7942;
 RA Holtman C.K., Sandoval P., Chen Y., Socias T., Mohler B.J.,
 RA McMurtry S., Gonzalez A., Salinas I., Golden S.S., Youderian P.;
 RT "Synechococcus elongatus PCC7942 cosmid 793."
 RL Submitted (JUN-2002) to the EMBL/GenBank/DDb databases.
 DR EMBL: AY120853; AAM82700.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 821 AA; 91706 MW; D5221A0DD0BD0BCC CRC64;

Query Match 32.3%; Score 54; DB 2; Length 821;
 Best Local Similarity 47.8%; Pred. No. 16;
 Matches 11; Conservative 2; Mismatches 10; Indels 0; Gaps 0;
 QY 1 PMSKCYQWQRRMKCAPSITC 23
 Db 517 PDGSOALMORRALTLTAKPTC 539

RESULT 14
 O97490 PRELIMINARY; PRT; 736 AA.
 AC O97490;
 DT 01-MAY-1998 (TREMBlrel. 10, Created)
 DT 01-MAY-1998 (TREMBlrel. 10, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE P97.
 OS *Oryctolagus cuniculus* (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=98451505; Pubmed=9780225;
 RA Kawamoto T., Pan H., Yan W., Ishida H., Usui E., Oda R., Nakamasu K.,
 RA Noshiro M., Kawashima-Ohya Y., Fujii M., Shintani H., Okada Y.,
 RA Kato Y.;
 RT "Expression of membrane-bound transferrin-like protein p97 on the cell
 surface of chondrocyte."
 RL Eur. J. Biochem. 256:503-509(1998).
 CC -I- FUNCTION: TRANSFERRIN ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
 CC -I- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC EMBL: AB010995; BAA33956.1; -.
 DR HSSP: P19134; ITPD.
 DR InterPro: IPR001064; Cysfallin.
 DR InterPro: IPR001156; Transferrin.
 DR Pfam: PF00405; Transferrin; 2.
 DR PRINTS: PR00422; TRANSFERRIN.
 DR SMART: SM00094; TR_FER; 2.
 DR PROSITE: PS00225; CRYSTALLIN BETAGAMMA; 1.
 DR PROSITE: PS00205; TRANSFERRIN 1; 2.
 DR PROSITE: PS00206; TRANSFERRIN 2; 1.
 DR PROSITE: PS00207; TRANSFERRIN 3; 1.
 DR GlycoProtein: Iron transport; Metal-binding; Transport.
 KW SEQUENCE 736 AA; 80169 MW; F389D9FBA6AC90FC CRC64;

Query Match 32.6%; Score 53.5; DB 6; Length 736;
 Best Local Similarity 36.7%; Pred. No. 17;
 Matches 11; Conservative 6; Mismatches 12; Indels 1; Gaps 1;

QY 1 PMSKCYQWQRRMKG-APSITCIRRTSA 29
 Db 31 PEQKCEDMSQAFRBAQLOPALLCVQGTSA 60

RESULT 15
 O77558 PRELIMINARY; PRT; 48 AA.
 ID O77558
 AC O77558;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Lactoferrin (Fragment).
 OS *Sus scrofa* (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 OX NCBI_TaxID=9923;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=98332734; Pubmed=9666128;
 RA Wang S.R., Lin J., Cheng I.C., Lin T.Y.;
 RT "Characterization and functional analysis of the porcine lactoferrin
 RT gene promoter."
 RL Gene 215:203-212 (1998).
 DR EMBL: AF044256; AAC34369.1; -.
 DR InterPro: IPR001156; Transferrin.
 DR Pfam: PF00405; Transferrin; 1.
 FT NON_TER 48
 SQ SEQUENCE 48 AA; 5569 MW; 28403BETDE144D78 CRC64;

Query Match 32.3%; Score 53; DB 6; Length 48;
 Best Local Similarity 57.1%; Pred. No. 1.3;
 Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 2 PMSKCYQWQRRMK 15
 Db 35 EYSKCRQWQSKIRR 48

Search completed: July 30, 2003, 16:29:25
 Job time : 51.1964 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:10:10 ; Search time 9.49405 Seconds

(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-5

Perfect score: 164
Sequence: 1 PEMSKCYQWQRMKLGAPSTICIRRTSA 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	164	100.0	708	1 TRFL_CAPHI	Q29477 capra hircu
2	124	75.6	708	1 TRFL_BUBBU	Q27698 bubalus bub
3	122	74.4	708	1 TRFL_BOVIN	P24627 bos taurus
4	104	63.4	708	1 TRFL_CAMDR	Q9tun0 camelus dro
5	100	61.0	711	1 TRFL_HUMAN	P02788 homo sapien
6	84	51.2	707	1 TRFL_MOUSE	P08071 mus musculu
7	83	50.6	695	1 TRFL_HORSE	O7811 equus cabal
8	66	40.2	704	1 TRFL_PIG	P14632 sus scrofa
9	58.5	35.7	704	1 ICA_PIG	Q29545 sus scrofa
10	58	35.4	271	1 MGP_ECOLI	Q84b99 escherichia
11	58	35.4	271	1 MGP_ECOLI	P76329 escherichia
12	58	35.4	271	1 MGP_SALTY	Q8xgel salmonella
13	57.5	35.1	738	1 TRFL_HUMAN	P08582 homo sapien
14	52	31.7	198	1 PEMT_MOUSE	O61907 mus musculu
15	52	31.7	271	1 MGP_ECOLI	P59286 escherichia
16	51.5	31.4	1795	1 ESMI_HUMAN	O14674 homo sapien
17	51	31.1	198	1 PEMT_RAT	Q08388 rattus norv
18	50	30.5	642	1 TRFL_HUMAN	O87412 homo sapien
19	50	30.5	706	1 TRFL_HORSE	P27425 equus cabal
20	50	30.5	1207	1 TRFL_HUMAN	O8e55 homo sapien
21	48	29.3	430	1 SYH_CHLBN	Q9v7p1 chlamydia p
22	47.5	29.0	450	1 PEMT_RAT	O9vwr7 rattus norv
23	47	28.7	198	1 PEMT_HUMAN	O12931 homo sapien
24	47	28.7	704	1 TRFL_HUMAN	O12931 homo sapien
25	47	28.7	2671	1 TRFL_HUMAN	O14573 homo sapien
26	46.5	28.4	695	1 TRFL_RABIT	P19134 onycholagus
27	46	28.0	63	1 BD01_MOUSE	Q9vcl0 mus musculu
28	46	28.0	413	1 YEIT_SALTY	O8e54 salmonella
29	46	28.0	413	1 YEIT_SALTY	O8e54 salmonella
30	46	28.0	488	1 VE2_HPV49	O8xnl8 salmonella
31	46	28.0	647	1 TOF3_VIBCH	O9qgfs vibrio chol
32	46	28.0	1082	1 A3B2_HUMAN	O13367 homo sapien
33	46	28.0	1210	1 AT19_MOUSE	P59509 mus musculu

34	45.5	27.7	641	1 SCAB_RABIT	O97742 onycholagus
35	45.5	27.7	698	1 TRFL_HUMAN	P02787 homo sapien
36	45.5	27.7	967	1 ATSI_HUMAN	Q9uh18 homo sapien
37	45	27.4	199	1 NHAA_PSECL	P27764 pseudomonas
38	45	27.4	310	1 TFE2H_METTH	O26971 pseudomonas
39	45	27.4	368	1 HA14_MOUSE	P14427 mus musculu
40	45	27.4	1385	1 RRP0_P1AMV	O07518 plantago as
41	45	27.4	1490	1 CRK7_HUMAN	O9vry4 homo sapien
42	44.5	27.1	937	1 SYL_METTH	O27552 methanobact
43	44.5	27.1	1793	1 YCF1_LOTUA	Q9bub6 lotus japon
44	44	26.8	142	1 RRP0_OENEB	P31843 oenothera b
45	44	26.8	236	1 YIHL_ECOLI	P32133 escherichia

ALIGNMENTS

```

RESULT 1
TRFL_CAPHI          STANDARD;          PRT;          708 AA.
AC      Q29477; Q29479;
DT      15-DEC-1998 (Rel. 37, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DB      Lactotransferrin precursor (Lactoferrin).
GN      LTF.
OS      Capra hircus (Goat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC      Bovidae; Caprinae; Capra.
OX      NCBI_TaxID=9925;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Mammary gland;
RA      Lee T., Yu S., Kim S., Lee K., Yu D.;
RL      Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Mammary gland;
RA      MEDLINE=94380047; Pubmed=8093048;
RX      le Provost F., Mocart M., Guerin G., Martin P.;
RT      "Characterization of the goat lactoferrin cDNA. Assignment of the
RT      relevant locus to bovine U12 synteny group.";
RL      Biochem. Biophys. Res. Commun. 203:1324-1332(1994).
CC      -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC      CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC      OF AN ANION. USUALLY BICARBONATE.
CC      -1- SUBUNIT: Monomer (By similarity).
CC      -1- SUBCELLULAR LOCATION: Secreted.
CC      -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC      -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U53857; AAA97958.1; -
CC      EMBL: X78902; CAAS5517.1; -
CC      HSSP: O77698; ICE2.
CC      InterPro: IPR001156; Transferrin.
CC      Pfam: PF00405; transferrin; 2.
CC      PRINTS: PF00422; TRANSFERRIN.
CC      SMART: SM00094; TR_FER; 2.
CC      PROSITE: PS00205; TRANSFERRIN_1; 2.
CC      PROSITE: PS00206; TRANSFERRIN_2; 2.
CC      PROSITE: PS00207; TRANSFERRIN_3; 2.
CC      Transport, Iron transport; Glycoprotein; Metal-binding; Repeat;
CC      Signal.
FT      SIGNAL.          1          19          BY SIMILARITY.

```

```

FT CHAIN 20 708 LACTOTRANSFERRIN.
FT REPEAT 20 363 1.
FT REPEAT 364 708 2.
FT DISULFID 28 64 BY SIMILARITY.
FT DISULFID 38 55 BY SIMILARITY.
FT DISULFID 134 217 BY SIMILARITY.
FT DISULFID 176 192 BY SIMILARITY.
FT DISULFID 189 200 BY SIMILARITY.
FT DISULFID 250 264 BY SIMILARITY.
FT DISULFID 367 399 BY SIMILARITY.
FT DISULFID 377 390 BY SIMILARITY.
FT DISULFID 424 703 BY SIMILARITY.
FT DISULFID 444 666 BY SIMILARITY.
FT DISULFID 476 551 BY SIMILARITY.
FT DISULFID 500 694 BY SIMILARITY.
FT DISULFID 510 524 BY SIMILARITY.
FT DISULFID 521 534 BY SIMILARITY.
FT DISULFID 592 606 BY SIMILARITY.
FT DISULFID 644 649 BY SIMILARITY.
FT METAL 79 79 IRON 1 (BY SIMILARITY).
FT METAL 111 111 IRON 1 (BY SIMILARITY).
FT METAL 211 211 IRON 1 (BY SIMILARITY).
FT METAL 272 272 IRON 1 (BY SIMILARITY).
FT METAL 414 414 IRON 2 (BY SIMILARITY).
FT METAL 452 452 IRON 2 (BY SIMILARITY).
FT METAL 545 545 IRON 2 (BY SIMILARITY).
FT METAL 614 614 IRON 2 (BY SIMILARITY).
FT BINDING 140 140 ANION (BY SIMILARITY).
FT CARBOHYD 482 482 ANION (BY SIMILARITY).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 56 56 I -> V (IN REF. 2).
FT CONFLICT 88 88 L -> R (IN REF. 2).
FT CONFLICT 124 124 Q -> K (IN REF. 2).
FT CONFLICT 154 154 F -> P (IN REF. 2).
FT CONFLICT 304 304 S -> R (IN REF. 2).
FT CONFLICT 414 414 D -> G (IN REF. 2).
SQ SEQUENCE 708 AA; 77358 MW; P2EDA3C83539960D CRC64;

```

Query Match 100.0%; Score 164; DB 1; Length 708;
 Best Local Similarity 100.0%; Pred No. 1.4e-16;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 PWSKCYOMORRMKLGAPSIICIRRTSA 29
    |||||
DB 33 PWSKCYOMORRMKLGAPSIICIRRTSA 61

```

RESULT 2

```

ID TRFL_BUBBU STANDARD; PRT; 708 AA.
AC 07/698;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lactotransferrin precursor (lactoferrin).
GN LTF.
OS Bubalus bubalis (Domestic water buffalo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bubalus.
OX NCBI_TaxID=89462;
RN [1]
RP SEQUENCE FROM N. A.
RA Paramesivam M., Thattaiyach B.D., Kumar A., Srinivasan A.,
RL "cDNA sequence of Buffalo lactoferrin.",
RN [2]
RP Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).

```

```

RX MEDLINE=20003130; PubMed=10531476;
RA Karthikeyan S., Paramesivam M., Yadav S., Srinivasan A., Singh T.P.;
RT "Structure of buffalo lactoferrin at 2.5-A resolution using crystals
RT grown at 303 K shows different orientations of the N and C lobes.";
RL Acta Crystallogr. D 55:1805-1813 (1999).
CC -1- FUNCTION: TRANSFERRIN ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION. USUALLY BICARBONATE.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ005203; CA06441.1; -.
DR PDB; 1CE2; 19-MAR-99.
DR PDB; 1BIY; 13-JAN-99.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SM00944; TR_FER; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
DR Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
KM Signal; 3D-structure.
FT CHAIN 1 19 LACTOTRANSFERRIN.
FT DISULFID 28 708
FT DISULFID 38 64
FT DISULFID 134 217
FT DISULFID 176 192
FT DISULFID 179 202
FT DISULFID 189 200
FT DISULFID 250 264
FT DISULFID 367 399
FT DISULFID 377 390
FT DISULFID 424 703
FT DISULFID 444 666
FT DISULFID 476 551
FT DISULFID 500 694
FT DISULFID 510 524
FT DISULFID 521 534
FT DISULFID 592 606
FT DISULFID 644 649
FT METAL 79 79 IRON 1.
FT METAL 111 111 IRON 1.
FT METAL 211 211 IRON 1.
FT METAL 272 272 IRON 1.
FT METAL 414 414 IRON 2.
FT METAL 452 452 IRON 2.
FT METAL 545 545 IRON 2.
FT METAL 614 614 IRON 2.
FT BINDING 140 140 ANION (POTENTIAL).
FT CARBOHYD 482 482 ANION (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 21 22
FT TURN 22 22
FT STAND 25 25
FT HELIX 32 46
FT TURN 47 47
FT STAND 53 57
FT HELIX 61 69
FT TURN 70 71

```



```

FT STRAND 75 78
PT HELIX 80 87
PT TURN 89 91
PT STRAND 93 102
PT STRAND 107 108
PT STRAND 110 118
PT TURN 119 120
PT HELIX 125 127
PT TURN 129 130
PT STRAND 132 135
PT TURN 138 139
PT TURN 141 144
PT HELIX 145 150
PT TURN 151 151
PT HELIX 152 155
PT TURN 159 161
PT HELIX 164 169
PT TURN 170 171
PT STRAND 174 176
PT TURN 178 179
PT TURN 182 184
PT HELIX 186 189
PT TURN 190 191
PT HELIX 196 198
PT TURN 199 200
PT TURN 204 205
PT TURN 207 208
PT HELIX 211 219
PT TURN 220 221
PT STRAND 225 229
PT TURN 230 231
PT HELIX 232 236
PT HELIX 240 243
PT TURN 244 245
PT STRAND 246 249
PT TURN 251 252
PT STRAND 255 257
PT HELIX 258 260
PT TURN 261 263
PT STRAND 267 270
PT STRAND 273 277
PT HELIX 283 297
PT TURN 299 300
PT TURN 307 308
PT TURN 312 313
PT TURN 321 322
PT STRAND 325 328
PT TURN 331 332
PT HELIX 335 351
PT HELIX 354 369
PT STRAND 369 384
PT HELIX 371 384
PT TURN 385 385
PT STRAND 388 393
PT HELIX 396 404
PT TURN 405 406
PT STRAND 410 413
PT HELIX 415 422
PT TURN 423 425
PT STRAND 426 434
PT TURN 438 441
PT TURN 444 446
PT STRAND 452 459
PT TURN 460 461
PT TURN 463 464
PT HELIX 467 469
PT TURN 471 472
PT STRAND 474 477
PT TURN 480 481
PT TURN 483 486
PT HELIX 487 497
PT TURN 503 505
PT STRAND 508 510

FT TURN 512 513
PT TURN 516 517
PT TURN 519 520
PT TURN 522 523
PT TURN 528 529
PT TURN 533 534
PT TURN 538 539
PT TURN 541 542
PT HELIX 544 553
PT TURN 554 555
PT STRAND 559 563
PT HELIX 564 569
PT TURN 570 570
PT HELIX 578 581
PT TURN 582 582
PT TURN 585 586
PT STRAND 588 591
PT TURN 593 594
PT STRAND 597 599
PT TURN 600 601
PT TURN 603 605
PT STRAND 609 611
PT STRAND 615 619
PT HELIX 620 637
PT TURN 639 640
PT TURN 642 647
PT TURN 650 651
PT TURN 654 655
PT TURN 662 663
PT STRAND 664 669
PT HELIX 676 680
PT HELIX 682 692
PT TURN 693 694
PT HELIX 698 705

Query Match Score 124; DB 1; Length 708;
Best Local Similarity 76.9%; Pred. No. 1.2e-10;
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PEMSKCYQOMRMRKLGAPSTICIR 26
Db 33 PEMKCHRMQWRMKKLGAPSTICVR 58

RESULT 3
TRFL_BOVIN STANDARD; PRT; 708 AA.
ID TRFL_BOVIN
AC P24627; Q29629; Q9MZY3;
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactoferrin precursor (Lactoferrin) [Contains: Lactoferricin B
DE (LpCIN B)].
GN LTF.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_Taxid=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Submaxillary gland;
RX MEDLINE=9116050; PubMed=2001696;
RA Pierce A., Colavizza D., Benaisa M., Maes P., Tatar A.,
RA Montreuil J., Spik G.;
RT "Molecular cloning and sequence analysis of bovine lactoferrin.";
RL Eur. J. Biochem. 196;177-184(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92028986; PubMed=1718281;
RA Goodman R.E., Schanbacher F.L.;
RT "Bovine lactoferrin mRNA: sequence, analysis, and expression in the
RT mammary gland.";
```

RL Biochem. Biophys. Res. Commun. 180:75-84(1991).

RL [3] SEQUENCE FROM N.A.

RP TISSUE=Lung;

RA Tsang T.C., Burns D.K., Wang F., Pan Y.C.E., Schmitz A.M., Stern D.;

RT "Cloning of a 80-kD advanced glycosylation end product (AGE) binding

RL protein from bovine lung."

RL FASEB J. 6:233-233(1991).

RP [4] SEQUENCE FROM N.A.

RC TISSUE=Blood, and Mammary gland;

RX MEDLINE=94266164; PubMed=206385;

RA Seyfert H.-M., Juckoritz A., Interthal H., Koczan D., Hobom G.;

RT "Structure of the bovine lactoferrin-encoding gene and its promoter."

RL Gene 143:265-269(1994).

RL [5] SEQUENCE FROM N.A.

RA Nakamura I., Shimazaki K., Yagi Y., Matanabe A.;

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

RN [6] SEQUENCE OF 20-59.

RX MEDLINE=90031466; PubMed=2805645;

RA Rejman J.J., Hegarty H.M., Hurley W.L.;

RT "Purification and characterization of bovine lactoferrin from

RT secretions of the involuting mammary gland: identification of

RT multiple molecular weight forms."

RL Comp. Biochem. Physiol. 93B:929-934(1989).

RN [7] X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).

RP MEDLINE=9606267; PubMed=9398529;

RX Moore S.A., Anderson B.F., Groom C.R., Haridas M., Baker E.N.;

RA "Three-dimensional structure of ferric bovine lactoferrin at 2.8-A

RT resolution."

RL J. Mol. Biol. 274:222-236(1997).

RN [8] STRUCTURE BY NMR OF 36-60.

RX MEDLINE=9819007; PubMed=9521752;

RA Hwang P.M., Zhou N., Shan X., Arrowsmith C.H., Vogel H.J.;

RT "Three-dimensional solution structure of lactoferricin B, an

RT antimicrobial peptide derived from bovine lactoferrin."

RL Biochemistry 37:4288-4298(1998).

CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH

CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING

CC OF AN ANION, USUALLY BICARBONATE.

CC -1- FUNCTION: LACTOFERRICIN B IS AN ANTIMICROBIAL PEPTIDE.

CC -1- SUBUNIT: Monomer.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.

CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL, X57084; CAA40366.1; -

DR EMBL, M63502; AAA30617.1; -

DR EMBL, L08604; AAA30609.1; -

DR EMBL, L19993; AAA21722.1; -

DR EMBL, L19982; AAA21722.1; JOINED.

DR EMBL, L19983; AAA21722.1; JOINED.

DR EMBL, L19984; AAA21722.1; JOINED.

DR EMBL, L19985; AAA21722.1; JOINED.

DR EMBL, L19986; AAA21722.1; JOINED.

DR EMBL, L19988; AAA21722.1; JOINED.

DR EMBL, L19989; AAA21722.1; JOINED.

DR EMBL, L19990; AAA21722.1; JOINED.

DR EMBL, L19991; AAA21722.1; JOINED.

DR EMBL, L19992; AAA21722.1; JOINED.

DR EMBL, AB046664; BAB03470.1; -

DR PIR; I45919; TFRHL.

DR PDB; 1BLF; 03-DEC-97.

DR PDB; 1LFC; 18-NOV-98.

DR GlycoSiteDB; P24627; -

DR InterPro; IPR001156; Transferrin.

DR Pfam; PF00405; transferrin; 2.

DR PRINTS; PRO0422; TRANSFERRIN.

DR SMART; SM00094; TR_FER; 2.

DR PROSITE; PS00205; TRANSFERRIN_1; 2.

DR PROSITE; PS00206; TRANSFERRIN_2.

DR PROSITE; PS00207; TRANSFERRIN_3; 2.

DR Transprot; Iron transport; glycoprotein; Metal-binding; Repeat;

KW Signal; Antithetic; 3D-structure.

FT SIGNAL 1 19

FT CHAIN 20 708

FT PEPTIDE 36 60

FT REPEAT 20 363

FT REPEAT 364 708

FT DISULFID 28 64

FT DISULFID 38 55

FT DISULFID 134 217

FT DISULFID 176 192

FT DISULFID 189 200

FT DISULFID 250 264

FT DISULFID 367 399

FT DISULFID 377 390

FT DISULFID 424 703

FT DISULFID 444 666

FT DISULFID 476 551

FT DISULFID 500 694

FT DISULFID 510 524

FT DISULFID 521 534

FT DISULFID 592 606

FT DISULFID 644 649

FT METAL 79 79

FT METAL 111 111

FT METAL 211 211

FT METAL 272 272

FT METAL 414 414

FT METAL 452 452

FT METAL 545 545

FT METAL 614 614

FT BINDING 140 140

FT BINDING 482 482

FT CARBOHYD 252 252

FT CARBOHYD 387 387

FT CARBOHYD 495 495

FT CARBOHYD 564 564

FT CONFLICT 63 63

FT CONFLICT 66 67

FT CONFLICT 145 145

FT CONFLICT 164 165

FT CONFLICT 264 264

FT CONFLICT 273 273

FT CONFLICT 281 281

FT CONFLICT 291 291

FT CONFLICT 297 297

FT CONFLICT 340 340

FT CONFLICT 418 418

FT CONFLICT 439 439

FT CONFLICT 459 459

FT CONFLICT 514 514

FT CONFLICT 632 632

FT STRAND 29 29

FT HELIX 32 45

FT HELIX 46 48

FT STRAND 53 57

FT HELIX 61 69

FT TURN 70 71

FT STRAND 75 78

RA -> A (IN REF. 4).

RA -> PG (IN REF. 2).

I -> V (IN REF. 1 AND 4).

LO -> PP (IN REF. 1).

C -> Y (IN REF. 4).

A -> P (IN REF. 4).

G -> A (IN REF. 4).

S -> R (IN REF. 4).

P -> S (IN REF. 2).

G -> A (IN REF. 1).

I -> V (IN REF. 1).

H -> Y (IN REF. 1).

K -> R (IN REF. 1).

A -> R (IN REF. 1).

H -> R (IN REF. 5).

```

FT HELIX 80 87
FT TURN 89 91
FT STRAND 93 102
FT STRAND 107 108
FT STRAND 110 118
FT TURN 126 127
FT TURN 129 130
FT STRAND 133 135
FT TURN 138 139
FT TURN 141 144
FT HELIX 145 155
FT TURN 156 156
FT TURN 159 161
FT HELIX 164 169
FT TURN 170 171
FT STRAND 175 176
FT TURN 178 179
FT TURN 182 184
FT HELIX 186 188
FT TURN 190 191
FT TURN 196 197
FT TURN 199 200
FT TURN 204 205
FT TURN 207 208
FT HELIX 210 219
FT TURN 220 221
FT STRAND 225 223

```

Query Match 74.4%; Score 122; DB 1; Length 708;
 Best Local Similarity 72.4%; Pred. No. 2.3e-10;
 Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

QY 1 PEMSKCYOMORMRKLGAPSTICIRRTSA 29
    ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 33 PEMSKCYOMORMRKLGAPSTICIRRTSA 61

```

RESULT 4
 TRFL_CAMDR STANDARD; PRT; 708 AA.

```

AC Q9TUM0; O9M255;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lactoferrin precursor (Lactoferrin).
GN LTF.
OS Camelus dromedarius (Dromedary) (Arabian camel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9938;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Small; TISSUE=Lactating mammary gland;
RA Kappeier S.R.; Ackermann M.; Farah Z.; Puhon Z.;
RT "Sequence analysis of camel (Camelus dromedarius) lactoferrin.";
RL Int. Dairy J. 9:481-486 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Paramaivam M.; Srinivasan A.; Singh R.; Sahani M.S.; Singh T.P.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: TRANSFERRING ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC OF AN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.

```

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isdb.ch/announce/>
 CC or send an email to license@isdb.ch).

```

CC -----
DR EMBL; AJ31674; CAB53387.1; -.
DR EMBL; AF165879; AAF82241.1; -.
DR PDB; 1DTZ; 20-JUN-01.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SM00394; TR_FER; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
KW Signal; Iron transport; Glycoprotein; Metal-binding; Repeat;
KW Signal; 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 708
FT REPEAT 20 363
FT REPEAT 364 708
FT DISULFID 28 64
FT DISULFID 38 55
FT DISULFID 134 217
FT DISULFID 176 192
FT DISULFID 189 200
FT DISULFID 250 264
FT DISULFID 367 399
FT DISULFID 377 390
FT DISULFID 424 703
FT DISULFID 444 666
FT DISULFID 476 551
FT DISULFID 500 694
FT DISULFID 510 524
FT DISULFID 521 534
FT DISULFID 592 606
FT DISULFID 644 649
FT METAL 79 79
FT METAL 111 111
FT METAL 211 211
FT METAL 272 272
FT METAL 414 414
FT METAL 452 452
FT METAL 545 545
FT METAL 614 614
FT BINDING 140 140
FT BINDING 482 482
FT CARBOHYD 252 252
FT CARBOHYD 385 385
FT CARBOHYD 537 537
FT CARBOHYD 594 594
FT CONFLICT 261 261
FT CONFLICT 304 304
FT CONFLICT 330 330
FT CONFLICT 492 494
FT CONFLICT 506 506
FT CONFLICT 609 609
FT CONFLICT 642 642
SQ SEQUENCE 708 AA; 77211 MW; 0B0C175A0B69430 CXC64;

```

Query Match 63.4%; Score 104; DB 1; Length 708;
 Best Local Similarity 63.0%; Pred. No. 1.1e-07;
 Matches 17; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

```

QY 2 EMSKCYOMORMRKLGAPSTICIRRTS 28
    ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34 EMSKCYOMORMRKLGAPSTICIRRTS 60

```

RESULT 5
 TRFL_HUMAN STANDARD; PRT; 711 AA.
 AC P02788; C00756; Q16780; Q16785; Q16786; Q16789; Q96K24; Q96K25;
 AC Q9H123;

DT 21-JUL-1986 (Rel. 01, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Lactoferrin precursor (lactoferrin) [Contains: Lactoferrroxin A;
Lactoferrroxin B; Lactoferrroxin C].

GN LTP OR LF.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eubacteria; Primates; Carnivora; Homiidae; Homo.

OC NCBI_Taxid=9606;

OK

RN

RP SEQUENCE FROM N.A.

RC TISSUE=Mammary gland;

RX MEDLINE=90384839; PubMed=2402455;

RA Rey M.W., Woloshuk S.L., de Boer H.A., Pieper F.R.;

RT "Complete nucleotide sequence of human mammary gland lactoferrin.";

RL Nucleic Acids Res. 18:5288-5288(1990).

RN

RP SEQUENCE FROM N.A.

RC TISSUE=Mammary gland;

RA Cho Y.Y.;

RT Thesis (1994). Genetic Engineering Research Institute / Taejeon, Korea.

RN

RP SEQUENCE FROM N.A.

RC TISSUE=Mammary gland;

RA Conneely O.M.;

RT Submitted (May-1992) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC TISSUE=Mammary gland;

RA Liang Q., Jimenez-Flores R., Richardson T.;

RT "Molecular cloning and sequence analysis of human lactoferrin.";

RL Submitted (DEC-1991) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC TISSUE=Bone marrow;

RA Wei X., Han J., Rado T.A.;

RT "Human neutrophil lactoferrin coding and 5' flanking region DNA
sequences.";

RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC TISSUE=Mammary gland;

RA TISSUE=Mammary gland;

RA Cheng H., Chen X., Huan L.;

RT "CDNA cloning and sequence analysis of human lactoferrin.";

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC TISSUE=Prostate;

RX MEDLINE=2238257; PubMed=12477932;

RA Strassberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klauser R.D., Collins F.S., Wagner L., Spemmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Hsieh F.,
Datchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loguella N.A., Peters G.J., Abramson R.D., Mullan S.J.,
Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Guntatone P.H.,
Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Rahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Small D.E.,
Schnerch A., Schin J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN

RP SEQUENCE OF 3-711 FROM N.A.

RC TISSUE=Mammary gland;

RX MEDLINE=90326549; PubMed=2374734;

RN

RA Powell M.J., Ogden J.E.;

RT "Nucleotide sequence of human lactoferrin cDNA.";

RL Nucleic Acids Res. 18:4013-4013(1990).

RN

RP SEQUENCE OF 20-711.

RX MEDLINE=8507667; PubMed=6510420;

RA Metz-Boutigue M.-H., Jolles J., Mazurier J., Schoentgen F.,
Legrand D., Spik G., Montreuil J., Jolles P.;

RT "Human lactoferrin: amino acid sequence and structural
comparisons with other transferrins.";

RL Eur. J. Biochem. 145:659-666(1984).

RN

RP PRELIMINARY SEQUENCE OF 20-73; 134-171; 257-278 AND 347-530.

RX MEDLINE=82046817; PubMed=6794640;

RA Metz-Boutigue M.-H., Mazurier J., Jolles J., Spik G., Montreuil J.,
Jolles P.;

RT "The present state of the human lactoferrin sequence. Study and
alignment of the cyanogen bromide fragments and characterization of
N- and C-terminal domains.";

RL Biochim. Biophys. Acta 670:243-254(1981).

RN

RP SEQUENCE OF 609-711.

RX MEDLINE=82262043; PubMed=7049727;

RA Metz-Boutigue M.-H., Jolles J., Mazurier J., Spik G., Montreuil J.,
Jolles P.;

RT "An 88 amino acid long C-terminal sequence of human
lactoferrin.";

RL FEBS Lett. 142:107-110(1982).

RN

RP SEQUENCE OF 436-711 FROM N.A.

RX MEDLINE=88001031; PubMed=3477300;

RA Rado T.A., Wei X., Benz E.J. Jr.;

RT "Isolation of lactoferrin cDNA from a human myeloid library and
expression of mRNA during normal and leukemic myelopoiesis.";

RL Blood 70:989-993(1987).

RN

RP SEQUENCE OF 237-711 FROM N.A.

RA McCombie W.R., Wilson R., Chen E., Gibbs R., Zuo L., Johnson D.,
Nhan M., Parnell L., Dedhia N., Ansari A., Maris E., Schutz K.,
Gao J.L., la Bastide M., Kaplan N., Greco T., Touchman J., Muzny D.,
Chen C.N., Evans C., Fitzgerald M., See L.H., Tang M., Porcel B.M.,
Dragan V., Giacalone J., Pae A., Powell E., Solinsky K.A., Desilva U.,
Diaz-Perez S., Zhou X., Yu Y., Watanabe M., Doggett N., Garcia D.,
Segripant U.L.;

RT Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

RN

RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND REVISIONS.

RX MEDLINE=90064528; PubMed=2585506;

RA Anderson B.F., Baker H.M., Norris G.E., Rice D.W., Baker E.N.;

RT "Structure of human lactoferrin: crystallographic structure analysis
and refinement at 2.8-A resolution.";

RL J. Mol. Biol. 209:711-734(1989).

RN

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).

RA Haridas M., Anderson B.F., Baker E.N.;

RT "Structure of human diferric lactoferrin refined at 2.2-A
resolution.";

RL Acta Crystallogr. D 51:629-646(1995).

RN

RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 20-353.

RX MEDLINE=97156796; PubMed=9001186;

RA Nicholson H., Anderson B.F., Bland T., Shewry S.C., Tweedie J.M.,
Baker E.N.;

RT "Mutagenesis of the histidine ligand in human lactoferrin: iron
binding properties and crystal structure of the histidine-
253--methionine mutant.";

RL Biochemistry 36:341-346(1997).

RN

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).

RX MEDLINE=99190892; PubMed=10089347;

RA Sun X.L., Baker H.M., Shewry S.C., Jameson G.B., Baker E.N.;

RT "Structure of recombinant human lactoferrin expressed in Aspergillus
awamori.";

```

RL Acta Crystallogr. D 55:403-407(1999).
RN [18]
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RN MEDLINE=99192677; PubMed=10089508;
RX Jameson G.B., Anderson B.F., Norris G.E., Thomas D.H., Baker E.N.,
RT "Structure of human apolactoferrin at 2.0-A resolution. Refinement
RL Acta Crystallogr. D 54:1319-1335(1998).
RN [19]
RX CHARACTERIZATION OF LACTOFERROXINS.
RN MEDLINE=91166929; PubMed=1369293;
RX Tani F., Iio K., Chiba H., Yoshikawa M.;
RT "Isolation and characterization of opioïd antagonist peptides derived
RL from human lactoferrin."
RN Agric. Biol. Chem. 54:1803-1810(1990).
RN [20]
RX VARIANTS THR-30 AND ARG-48.
RN MEDLINE=99091914; PubMed=9873069;
RX Klintworth G.K., Sommer J.R., O'Brien G., Han L., Ahmed M.N.,
RA Qumsiyeh M.B., Lin P.-Y., Bacci S., Reddy M.K., Kanai A., Hotta Y.,
RA Sugar J., Kumaraswami Kaveli G., Munier F., Schorderet D.F.,
RA El Macti L., Iwata F., Kaiser-Kupfer M., Nagata M., Nakayasu K.,
RA Hejtmancik J.F., Teng C.T.;
RT "Familial subepithelial corneal amyloidosis (gelatinous drop-like
RL corneal dystrophy): exclusion of linkage to lactoferrin gene."
RN Mol. Vision 4:31-32(1998).
CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE.
CC -1- FUNCTION: LACTOFERROXINS A, B AND C HAVE OPIOID ANTAGONIST
CC ACTIVITY. LACTOFERROXIN A SHOWS PREFERENCE FOR MU-RECEPTORS, WHILE
CC LACTOFERROXIN B AND C HAVE SOMEWHAT HIGHER DEGREES OF PREFERENCE
CC FOR KAPPA-RECEPTORS THAN FOR MU-RECEPTORS.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53961; CA37914.1; -
DR EMBL; U07643; AAB60324.1; -
DR EMBL; M93150; AAA56159.1; -
DR EMBL; M83202; AAA59511.1; -
DR EMBL; M83205; AAA58656.1; -

```

```

Query Match 61.0%; Score 100; DB 1; Length 711;
Best Local Similarity 57.1%; Pred. No. 4.2e-07;
Matches 16; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
OY 1 PEMSKCYOMORMRKKGAPSTICIRRTS 28
DB 34 PEATKCFOMORMRKRGPPVSCIRKDS 61

```

```

RESULT 6
TRFL_MOUSE STANDARD; PRT; 707 AA.
AC P08071; P70690; O61799; O922P2;
DT 01-AUG-1988 (Rel. 08, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lactotransferrin precursor (lactoferrin).
GN LTF.
OS Mus musculus (Mouse).
OC Buxaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

OX NCBI_TaxID=10090;
RN [1]
RX SEQUENCE FROM N.A.
RN TISSUE=Uterus;
RX MEDLINE=87280033; PubMed=3611056;
RA Pentecost B.T., Teng C.T.;
RT "Lactotransferrin is the major estrogen inducible protein of mouse
RL uterine secretions."
RN J. Biol. Chem. 262:10134-10139(1987).
RN [2]
RX SEQUENCE FROM N.A.
RN TISSUE=Uterus;
RA Sumitani K.;
RN Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RX SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strauberg R.L., Feinold E.A., Gronse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carinini P., Prange C.,
RA Raha S.S., Iguellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Boeck S.A., McEwan P.J., McKernan K.U., Malek J.A., Gamaralle P.H.,
RA Richards S., Motley K.C., Hale S., Garcia A.M., Gay L.J., Huylk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.U., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywicki M.I., Skalski U., Smalins D.E.,
RA Scherich A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RX SEQUENCE OF 1-14 FROM N.A.
RX MEDLINE=92042099; PubMed=1939212;
RA Liu Y., Teng C.T.;
RT "Characterization of estrogen-responsive mouse lactoferrin promoter."
RL J. Biol. Chem. 266:21880-21885(1991).
CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U03298; AAA40525.1; -
DR EMBL; D88510; BA11633.1; -
DR EMBL; BC006904; AAH06904.1; -
DR EMBL; M74778; AAA39427.1; -
DR HSSP; P02788; ICB6.
DR MGD; MGI:96837; Ltf.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; Transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SM00094; TR_FER; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 1.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
DR Transport; Iron transport; Metal-binding; Repeat;

```

KM Signal. 1 19 BY SIMILARITY.
 FT SIGNAL 20 707 LACTOTRANSFERRIN.
 FT CHAIN 20 357 1.
 FT REPEAT 358 707 2.
 FT DISULFID 27 63 BY SIMILARITY.
 FT DISULFID 37 54 BY SIMILARITY.
 FT DISULFID 133 216 BY SIMILARITY.
 FT DISULFID 175 191 BY SIMILARITY.
 FT DISULFID 188 199 BY SIMILARITY.
 FT DISULFID 249 263 BY SIMILARITY.
 FT DISULFID 366 398 BY SIMILARITY.
 FT DISULFID 376 389 BY SIMILARITY.
 FT DISULFID 423 702 BY SIMILARITY.
 FT DISULFID 443 665 BY SIMILARITY.
 FT DISULFID 475 550 BY SIMILARITY.
 FT DISULFID 499 693 BY SIMILARITY.
 FT DISULFID 509 523 BY SIMILARITY.
 FT DISULFID 520 533 BY SIMILARITY.
 FT DISULFID 591 605 BY SIMILARITY.
 FT DISULFID 643 648 BY SIMILARITY.
 FT METAL 78 78 IRON 1 (BY SIMILARITY).
 FT METAL 110 110 IRON 1 (BY SIMILARITY).
 FT METAL 210 210 IRON 1 (BY SIMILARITY).
 FT METAL 271 271 IRON 1 (BY SIMILARITY).
 FT METAL 413 413 IRON 2 (BY SIMILARITY).
 FT METAL 451 451 IRON 2 (BY SIMILARITY).
 FT METAL 544 544 IRON 2 (BY SIMILARITY).
 FT METAL 613 613 IRON 2 (BY SIMILARITY).
 FT BINDING 139 139 ANTON (POTENTIAL).
 FT BINDING 481 481 ANTON (POTENTIAL).
 FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 1 2 R -> Q (IN REF. 2).
 FT CONFLICT 25 25 R -> Q (IN REF. 2).
 FT CONFLICT 82 82 S -> T (IN REF. 2).
 FT CONFLICT 359 359 S -> T (IN REF. 2).
 FT CONFLICT 382 382 A -> D (IN REF. 1).
 FT CONFLICT 449 449 E -> G (IN REF. 2).
 FT CONFLICT 629 629 L -> V (IN REF. 1).
 SQ SEQUENCE 707 AA; 77865 MM; F26AE0340AC19A8 CRC64;
 Query Match 51.2%; Score 84; DB 1; Length 707;
 Best Local Similarity: 44.4%; Pred. No. 9.8e-05;
 Matches 12; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

RA Sharma A.K., Paramesivam M., Srinivasan A., Yadav M.P., Singh T.P.;
 RT "Three-dimensional structure of mare deferetic lactoferrin at 2.6-A
 resolution.";
 RL J. Mol. Biol. 289:303-317(1999).
 CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION, USUALLY BICARBONATE.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -1- SIMILARITY: BELONGS TO TWO HOMOLOGOUS DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ010930; CAA09407.1; -.
 DR PDB; 1B1X; 02-DEC-98.
 DR PDB; 1B7U; 02-FEB-99.
 DR PDB; 1B7Z; 02-FEB-99.
 DR PDB; 1P98; 10-FEB-01.
 DR PDB; 1I68; 13-FEB-02.
 DR PDB; 1QJW; 14-JAN-00.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; Transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 1.
 DR Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 KM Signal; 3D-structure.
 FT SIGNAL 1 1
 FT CHAIN 7 66 LACTOTRANSFERRIN.
 FT REPEAT 7 350 1.
 FT REPEAT 351 695 2.
 FT DISULFID 15 51
 FT DISULFID 25 42
 FT DISULFID 121 204
 FT DISULFID 163 179
 FT DISULFID 166 189
 FT DISULFID 176 187
 FT DISULFID 237 251
 FT DISULFID 354 386
 FT DISULFID 364 377
 FT DISULFID 411 690
 FT DISULFID 431 653
 FT DISULFID 463 538
 FT DISULFID 487 681
 FT DISULFID 497 511
 FT DISULFID 508 521
 FT DISULFID 579 593
 FT DISULFID 631 636
 FT METAL 66 66
 FT METAL 98 98
 FT METAL 198 198
 FT METAL 259 259
 FT METAL 401 401
 FT METAL 439 439
 FT METAL 532 532
 FT METAL 601 601
 FT METAL 127 127
 FT BINDING 127 127 ANTON (BY SIMILARITY).
 FT BINDING 469 469 ANTON (BY SIMILARITY).
 FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 482 482 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 12 16
 FT HELIX 19 34

```

FT TURN 35 36
PT STRAND 40 44
FT HELIX 48 56
PT TURN 57 58
PT STRAND 62 65
FT HELIX 67 74
PT TURN 76 78
FT STRAND 80 88
PT STRAND 95 95
PT STRAND 97 105
PT TURN 106 107
FT HELIX 112 114
PT TURN 116 117
FT STRAND 119 122
PT TURN 125 126
FT TURN 128 131
FT HELIX 132 137
PT HELIX 139 142
PT TURN 147 148
FT HELIX 151 158
PT STRAND 161 163
PT TURN 165 166
PT TURN 169 171
FT HELIX 173 175
PT TURN 176 178
PT TURN 183 187
PT TURN 191 192
PT TURN 194 195
FT HELIX 197 206
PT TURN 207 208
FT STRAND 212 216
PT HELIX 219 223
PT HELIX 227 230
PT TURN 231 232
PT STRAND 233 237
PT TURN 238 240
PT STRAND 241 244
PT HELIX 245 247
PT TURN 248 250
PT STRAND 254 257
PT STRAND 260 264
PT HELIX 270 284
PT TURN 286 287
PT TURN 294 295
PT TURN 308 309
PT STRAND 312 315
PT HELIX 322 326
PT HELIX 328 336
PT TURN 337 338
PT STRAND 341 349
PT HELIX 351 356
PT HELIX 358 370
PT TURN 371 372
PT STRAND 375 380
PT HELIX 383 392
PT TURN 393 393
PT STRAND 397 400
PT HELIX 402 410
PT TURN 411 412
PT STRAND 414 421
PT HELIX 429 433
PT STRAND 439 446
PT HELIX 454 456
PT TURN 458 459
PT STRAND 462 464
PT TURN 467 468
PT TURN 470 473
PT HELIX 474 484
PT HELIX 489 491
PT TURN 492 492
PT STRAND 496 497
PT TURN 499 500
PT TURN 503 504

```

```

FT TURN 506 507
PT STRAND 508 508
PT TURN 515 516
PT TURN 520 521
PT TURN 525 526
FT TURN 528 529
FT HELIX 531 540
FT TURN 541 542
PT STRAND 546 550
PT HELIX 551 555
PT TURN 557 557
PT TURN 559 560
FT HELIX 565 568
PT TURN 569 569
FT TURN 572 573
PT STRAND 575 578
PT TURN 580 581
PT STRAND 584 586
FT HELIX 587 592
PT STRAND 596 598
PT STRAND 602 605
PT TURN 607 609
FT HELIX 610 624
PT TURN 626 627
PT TURN 629 630
PT HELIX 631 634
PT TURN 637 638
PT TURN 641 642
PT TURN 649 650
PT STRAND 651 655
PT TURN 658 659
PT HELIX 663 667
PT HELIX 669 679
PT TURN 680 681

```

Query Match 50.6%; Score 83; DB 1; Length 695;
 Best Local Similarity 50.0%; Pred. No. 0.00013;
 Matches 14; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

Qy 2 EWSKCYOMRRMRKLGAPSTICIRPTSA 29
 Db 21 EAAKCAKFORNMKRGPSVSCIRKTS 48

```

RESULT 8
ID TRFL_PIG STANDARD; PRT; 704 AA.
AC P14632; Q29557;
DT 01-APR-1990 (Rel. 14, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactoferrin precursor (Lactoferrin).
GN LTF.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92379101; PubMed=1511016;
RA Lydon J.P., O'Malley B.R., Saucedo O., Lee T., Headon D.R.,
RA Conneely O.M.;
RT "Nucleotide and primary amino acid sequence of porcine lactoferrin.";
RL Biochim. Biophys. Acta 1132:97-99(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92367939; PubMed=1503259;
RA Alexander L.J., Levine W.B., Teng C.T., Beattie C.W.;
RT "Cloning and sequencing of the porcine lactoferrin cDNA.";
RL Anim. Genet. 23:251-256(1992).
RN [3]
RP SEQUENCE OF 20-49.
RX MEDLINE=90105538; PubMed=2605266;

```


DT 28-FEB-2003 (Rel. 41, last sequence update)
 DT 15-SEP-2003 (Rel. 42, last annotation update)
 DE Putative mannoseyl-3-phosphoglycerate phosphatase (EC 3.1.3.70) (MGCP).
 GN YEDP OR STM1986 OR STY2193 OR T0892.
 OS Salmoneilla typhimurium, and
 OS Salmoneilla typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmoneilla.
 OX NCBI_TaxID=602, 601;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. typhimurium; STRAIN=LT2 / SSGS1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.,
 RT "Complete genome sequence of Salmoneilla enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856 (2001).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC SPECIES=S. typhi; STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parthill J., Dougan G., James K.D., Thomson N.R., Richard D., Main J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,
 RA Baker S., Bauman D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagsels K.,
 RA Krogh A., Larsen T.S., Leather S., Mount S., O'Garra P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.,
 RT "Complete genome sequence of a multiple drug resistant Salmoneilla
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852 (2001).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC SPECIES=S. typhi; STRAIN=TY2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liu S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyanni V., Schwartz D.C., Blatter F.R.,
 RT "Comparative genomics of Salmoneilla enterica serovar Typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337 (2003).
 CC -1- CATALYTIC ACTIVITY: 2(alpha-D-mannosyl)-3-phosphoglycerate + H(2)O.
 CC = 2(alpha-D-mannosyl)-D-glycerate + phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE HAD SUPERFAMILY. MGPB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE008788; AAL20896.1; -
 DR EMBL: AL627272; CAD05733.1; -
 DR EMBL: AE016837; AAO68570.1; -
 DR StGene; SG77777; YedP.
 DR HAMAP: MF_00617; -; 1.
 DR InterPro: IPR006379; HAD_SF_IIB.
 DR InterPro: IPR006381; HAD_SF_YedP.
 DR InterPro: IPR00634; Hydrolase.
 DR Pfam: PF00702; Hydrolase; 1.
 DR TIGRFAMs: TIGR01486; HAD-SF-IIB-YedP; 1.
 DR TIGRFAMs: TIGR01484; HAD-SF-IIB; 1.
 KW Hypothetical protein, Hydrolase; Complete proteome.
 SQ SEQUENCE 271 AA; 30925 MW; 46866C2E36B7008E CRC64;

Query Match

35.4%; Score 58; DB 1; Length 271;

Best Local Similarity 35.7%; Pred. No. 0.25;
 Matches 10; Conservative 4; Mismatches 14; Indels 0; Gaps 0;
 Qy 2 EWSKCYQWRMRKLAAPSITTCIRRTSA 29
 Db 25 EWCPAAPMLTRLHESGVPLTCSKTA 52
 RESULT 13
 TRFM_HUMAN
 ID TRFM_HUMAN STANDARD; PRT; 738 AA.
 AC P08582; Q9B022;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DE 15-SEP-2003 (Rel. 42, last annotation update)
 DE Melanocrasferitin precursor (Melanoma-associated antigen p97) (CD228
 DE antigen).
 GN MF12 OR MAP97.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Melanoma;
 RX MEDLINE=86149285; PubMed=2419904;
 RA Rose T.W., Plozman G.D., Teplow D.B., Dreyer W.J., Hellstroem K.E.,
 RA Brown J.P.,
 RT "Primary structure of the human melanoma-associated antigen p97
 RT (melanocrasferin) deduced from the mRNA sequence.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:1261-1265 (1986).
 RN [2]
 RN SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Skin and Uremia;
 RX MEDLINE=22389257; PubMed=12477932;
 RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shermen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopling R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carrinci P., Prange C.,
 RA Rahn S.S., Loguclano N.A., Peters G.J., Abraham R.D., Mullaly S.J.,
 RA Boeck S.A., McKean P.J., McKernan K.J., Malek U.A., Gamarate P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.W., Gay L.J., Huliy S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]
 RN GPI-ANCHOR.
 RX MEDLINE=94132080; PubMed=8300636;
 RA Food M.R., Rothenberger S., Gabathuler R., Haidl I.D., Reid G.,
 RA Vetterli W.A.,
 RT "Transport and expression in human melanomas of a transferrin-like
 RT glycosylphosphatidylinositol-anchored protein.";
 RL J. Biol. Chem. 269:3034-3040 (1994).
 RN [4]
 RN FUNCTION.
 RX MEDLINE=96016189; PubMed=7556058;
 RA Kennard M.L., Richardson D.R., Gabathuler R., Ponka P.,
 RA Vetterli W.A.,
 RT "A novel iron uptake mechanism mediated by GPI-anchored human p97.";
 RL EMBD J. 14:4178-4186 (1995).
 RN [5]
 RN IRON-BINDING.
 RX MEDLINE=92183868; PubMed=1544447;

FT	METAL	451	451	IRON 2 (BY SIMILARITY)
FT	METAL	556	556	IRON 2 (BY SIMILARITY)
FT	METAL	625	625	IRON 2 (BY SIMILARITY)
FT	BLINDING	136	136	ANTION (POTENTIAL)
FT	TRANSMEM	738	738	ANCHOR
FT	CARBOHYD	38	38	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	135	135	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	515	515	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	VARSPPLIC	238	302	GKTLPSKQALLSDPELCLCPGSRADYVEMOCHIAVPA HAYVYRDTGCGITFLINLEGORL -> ESRERDTWRSE EESBECFAHEARTTRSSAQAQMKVAPHRPODESNGEF /PrtId=VSP_006557. /PrtId=VSP_006558. Missing (in isoform 2) . /PrtId=VSP_006558.
SO	SEQUENCE	738 AA;	80241 MM;	6E6086E894D7B955 CRG64;
Qy	Query Match	35.1%	Score 57.5;	DB 1; Length 738;
Db	Best Local Similarity	43.3%	Pred. No. 0.85;	
	Matches 13;	Conservative 3;	Mismatches 13;	Indels 1; Gaps 1;
Db	1 PMSKCYQMORMRRLG-APSITCIRRTSA 29			
	31 PEGHKGNWSEAFREAGIQPILCVRTSA 60			
RESULT 14				
PEMT_MOUSE				
ID	PEMT_MOUSE	STANDARD;	PRT;	198 AA.
AC	061907;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	phosphatidylethanolamine N-methyltransferase (EC 2.1.1.17) (PEAMT)			
DE	(PEMT) (PEMT2).			
GN	PEMT OR PEMT2 OR PEMT2.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxId=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-129/J; TISSUE=Liver.			
RX	MEDLINE=97133127; PubMed=8978486;			
RA	Walkey C.U., Donohue L.R., Bronson R., Agellon L.B., Vance D.E.;			
RT	"Disruption of the murine gene encoding phosphatidylethanolamine			
RT	N-methyltransferase-2 gene.";			
RT	J. Lipid Res. 37:2341-2350(1996).			
RN	[2]			
RP	CHARACTERIZATION.			
RC	MEDLINE=98058738; PubMed=9371769;			
RX	Walkey C.U., Donohue L.R., Bronson R., Agellon L.B., Vance D.E.;			
RT	"Biochemical and evolutionary significance of phospholipid			
RT	methylation.";			
RT	J. Biol. Chem. 273:27043-27046(1998).			
CC	- FUNCTION: CATALYZE THREE SEQUENTIAL METHYLATION OF			
CC	PHOSPHATIDYLETHANOLAMINE (PE) BY ADOMET, THUS PRODUCING			
CC	PHOSPHATIDYLCHOLINE (PC).			
CC	- CATALYTIC ACTIVITY: S-adenosyl-L-methionine +			
CC	phosphatidylethanolamine = S-adenosyl-L-homocysteine +			
CC	phosphatidyl-N-methylethanolamine.			
CC	SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL). FOUND			
CC	IN MITOCHONDRIA AND IN ENDOPLASMIC RETICULUM.			
CC	- TISSUE SPECIFICITY: LIVER.			
CC	- SIMILARITY: BELONGS TO THE PEMT / PEM2 METHYLTRANSFERASE FAMILY.			

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U25051; AAA67686.1; JOINED.
 DR EMBL; U25046; AAA67686.1; JOINED.
 DR EMBL; U25047; AAA67686.1; JOINED.
 DR EMBL; U25048; AAA67686.1; JOINED.
 DR EMBL; U25049; AAA67686.1; JOINED.
 DR EMBL; U25050; AAA67686.1; JOINED.
 DR MGJ; MGJ:104535; Pamt.
 DR Pfam; PF004191; Pamt. 1.
 KW Phospholipid biosynthesis; Transferase; Methyltransferase;
 KW Transmembrane; Mitochondrion; Endoplasmic reticulum.
 FT INIT_Met 0 0 BY SIMILARITY.
 FT TRANSMEM 12 32 POTENTIAL.
 FT TRANSMEM 45 65 POTENTIAL.
 FT TRANSMEM 90 110 POTENTIAL.
 FT TRANSMEM 157 177 POTENTIAL.
 SQ SEQUENCE 198 AA; 22385 MW; 582B975A5173D72 CRC64;

Query Match 31.7%; Score 52; DB 1; Length 198;
 Best Local Similarity 36.0%; Pred. No. 1.4;
 Matches 9; Conservative 6; Mismatches 6; Indels 4; Gaps 1;

QY 3 WSKCYOWRRMRKLT---GAPSIYC 23
 DB 28 WNVVAFWEGRTKLSRAFGSPHLAC 52

RESULT 15
 MPGP_ECOL6
 ID MPGP_ECOL6 STANDARD; PRT; 271 AA.
 AC P59286;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative mannosyl-3-phosphoglycerate phosphatase (EC 3.1.3.70) (MPGP).
 GN YEDP OR C2373.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / ATCC 700928;
 RX MEDLINE=22386234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 RL -1- CATABOLIC ACTIVITY: 2(alpha-D-mannosyl)-3-phosphoglycerate + H(2)O
 CC = 2(alpha-D-mannosyl)-D-glycerate + phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE HAD SUPERFAMILY. MPGP FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE016762; AAN80832.1; -
 DR HAMAP; MF_00617; -; 1.

DR Pfam; PF00702; Hydrolase; 1.
 DR TIGRFAMs; TIGR01486; HAD-SF-IIb-YedP; 1.
 DR TIGRFAMs; TIGR01484; HAD-SF-IIb; 1.
 KW Hypothetical protein; Hydrolase; Complete proteome.
 SQ SEQUENCE 271 AA; 30515 MW; E6EB78E391470B2 CRC64;

Query Match 31.7%; Score 52; DB 1; Length 271;
 Best Local Similarity 32.1%; Pred. No. 2;
 Matches 9; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY 2 WSKCYOWRRMRKLTGAPSIYTCIRRTSA 29
 DB 25 DMQPAAPWLSRLHANIPLYLCSKTSIA 52

Search completed: July 30, 2003, 16:24:30
 Job time : 10.494 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:55 ; Search time 18.4702 Seconds
(without alignments)
150.994 Million cell updates/sec

Title: US-09-787-070-5

Perfect score: 164
Sequence: 1 PEMSKCYQWQRRMRKLGAPSTICIRRTSA 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	163	99.4	708	2 JC2323	lactoferrin - goat
2	122	74.4	708	1 TFBOL	lactoferrin p
3	100	61.0	711	1 TFBUL	lactoferrin p
4	91	55.5	33	2 S52107	lactoferrin - shee
5	84	51.2	707	1 A28438	lactoferrin precu
6	66	40.2	703	2 A45543	lactoferrin precu
7	58.5	35.7	704	2 I47228	carbonic anhydrase
8	58	35.4	271	2 B85813	hypothetical prote
9	58	35.4	271	2 E90965	hypothetical prote
10	58	35.4	271	2 G64959	hypothetical prote
11	58	35.4	271	2 AB0754	conserved hypochet
12	57.5	35.1	738	1 TFBUL	melanotransferrin
13	51	31.1	199	2 A47353	phosphatidylethano
14	51	31.1	489	2 C70655	probable monooxyme
15	50	30.5	706	2 S33761	transferrin precu
16	49	29.9	405	2 AB1461	B. subtilis yabB p
17	49	29.9	408	2 AC1098	B. subtilis yabB p
18	49	29.9	507	2 G87466	cryptophan halogen
19	49	29.9	1721	2 T21214	hypothetical prote
20	48.5	29.6	71	2 AH0033	hypothetical prote
21	48	29.3	430	2 D86573	histidyl tRNA synt
22	48	29.3	430	2 H72052	histidyl tRNA lig
23	47.5	29.0	423	2 F86313	P2115.8 protein -
24	47	28.7	58	2 S39410	noar protein - par
25	47	28.7	528	2 T21834	hypothetical prote
26	47	28.7	795	2 T21834	eceriferum3 (CER3)
27	47	28.7	822	2 AB2507	hypothetical prote
28	47	28.7	2671	2 A49873	insitol 1,4,5-tri
29	46.5	28.4	121	2 AH3147	hypothetical prote

30	46.5	28.4	694	1 TFRBP
31	46	28.0	413	2 AH0780
32	46	28.0	469	2 T26463
33	46	28.0	488	2 S36570
34	46	28.0	622	2 S61692
35	46	28.0	647	2 E82126
36	46	28.0	1082	2 T50650
37	45.5	27.7	384	2 E69050
38	45.5	27.7	550	2 T47158
39	45.5	27.7	620	2 T27008
40	45.5	27.7	638	1 TFRBP
41	45	27.4	200	1 A42725
42	45	27.4	284	2 S58650
43	45	27.4	310	2 C69218
44	45	27.4	368	2 I55961
45	45	27.4	395	2 T31578

ALIGNMENTS

RESULT 1

JC2323
lactoferrin - goat
C:Species: Capra aegagrus hircus (domestic goat)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 07-May-1995
C:Accession: JC2323
R:Le Provost, F.; Nocart, M.; Guerin, G.; Martin, P.
Biochem. Biophys. Res. Commun. 203, 1334-1332, 1994
A>Title: Characterization of the goat lactoferrin cDNA: assignment of the relevant locus
A:Reference number: JC2323; MUID:94380047; PMID:8093048
A:Accession: JC2323
A:Molecule type: mRNA
A:Residues: 1-708 <LEP>
C:Superfamily: transferrin; transferrin repeat homology
C:Keywords: duplication; glycoprotein
F:359-696/Domain: transferrin repeat homology <TRH2>
F:252,300,387,495,564/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match

Beat Local Similarity 99.4%; Score 163; DB 2; Length 708;
Matches 28; Conservative 1; Mismatches 0; Gaps 0;

QY 1 PEMSKCYQWQRRMRKLGAPSTICIRRTSA 29
DB 33 PEMSKCYQWQRRMRKLGAPSTICIRRTSA 61

RESULT 2

TFBOL
lactotransferrin precursor - bovine
N:Alternate names: Lactoferrin
C:Species: Bos primigenius taurus (cattle)
C:Date: 31-Mar-1992 #sequence_revision 21-Nov-1997 #text_change 11-May-2000
C:Accession: I45919; S14674; S14110; S18511; J10595; S13097; S18518; S13881; P10148; S21
R:Tsang, T.C.; Burns, D.K.; Wang, F.; Pan, Y.
PAGES 5, 6, 233, 1991
A>Title: Cloning of a 80-kD advanced glycosylation end product (AGE) binding protein frc
A:Reference number: I45919
A:Accession: I45919
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-708 <TSA>
A:Cross-references: GB:I08604; NID:g163269; PIDN:AAA30609.1; PID:g163270
R:Pierce, A.
submitted to the EMBL Data Library, November 1990
A:Reference number: S14674
A:Accession: S14674
A:Molecule type: mRNA
A:Residues: 1-144, 'V', 146-163, 'PP', 166-339, 'A', 341-438, 'Y', 440-513, 'R', 515-708 <PI1>
A:Cross-references: EMBL:X57084; NID:9505; PIDN:CAA40366.1; PID:9506
R:Pierce, A.; Colvitzza, D.; Benalissa, M.; Maes, P.; Tartar, A.; Montreuil, J.; Spik, G.
Eur. J. Biochem. 196, 177-184, 1991

A/Title: Molecular cloning and sequence analysis of bovine lactoferrin.
 A/Reference number: S14110; MUID:9160550; PMID:2001696
 A/Accession: S14110
 A/Molecule type: mRNA
 A/Residues: 3-144, 'V', 146-339, 'A', 341-438, 'Y', 440-513, 'R', 515-708 <PI2>
 A/Cross-references: EMBL:X57084
 A/Accession: S18517
 A/Molecule type: protein
 A/Residues: 20-35;82-114;148-163, 'P', 166-178, 'V', 'P', 183-190;205-212;230-239;304-339;59
 R/Goodman, R.E.; Schanbacher, F.L.
 Biochem. Biophys. Res. Commun. 180, 75-84, 1991
 A/Title: Bovine lactoferrin mRNA: sequence, analysis, and expression in the mammary gland
 A/Reference number: J70595; MUID:9202986; PMID:1718281
 A/Accession: J70595
 A/Molecule type: mRNA
 A/Residues: 1-65, 'PG', 68-296, 'S', 298-339, 'A', 341-708 <GOO>
 A/Cross-references: GB:M63502
 A/Note: the authors translated the codon CCG for residue 66 as Arg and TCT for residue 2
 R/Mead, P.E.; Tweedie, J.W.
 Nucleic Acids Res. 18, 7167, 1990
 A/Title: cDNA and protein sequence of bovine lactoferrin.
 A/Reference number: S13097; MUID:91088328; PMID:2263492
 A/Accession: S13097
 A/Molecule type: mRNA
 A/Residues: 28-33, 'DS', 36-38, 'P', 40-708 <MEA>
 A/Cross-references: EMBL:X54801
 A/Accession: S18518
 A/Molecule type: protein
 A/Residues: 20-47;59-66;132-139;256-277;278, 305-332;343-351;361-363;586;587-589;598-619
 R/Mead, P.E.
 submitted to the EMBL Data Library, October 1990
 A/Reference number: S13881
 A/Accession: S13881
 A/Molecule type: mRNA
 A/Residues: 28-38, 'P', 40-86, 'C', 88-708 <ME3>
 A/Cross-references: EMBL:X54801
 R/Reiman, J.U.; Hegarty, H.M.; Hurley, W.L.
 Comp. Biochem. Physiol. B 93, 929-934, 1989
 A/Title: Purification and characterization of bovine lactoferrin from secretions of the
 A/Reference number: P10148; MUID:90031466; PMID:2805645
 A/Accession: P10148
 A/Molecule type: protein
 A/Residues: 20-27, 'X', 29-37, 'X', 39-54, 'X', 56-59 <REJ>
 R/Bellamy, W.; Takase, M.; Yamuchi, K.; Wakabayashi, H.; Kawase, K.; Tomita, M.
 Biochim. Biophys. Acta 1121, 130-136, 1992
 A/Title: Identification of the bactericidal domain of lactoferrin.
 A/Reference number: S21756; MUID:92287941; PMID:1599334
 A/Accession: S21756
 A/Molecule type: protein
 A/Residues: 36-60 <BEL>
 R/Shimazaki, K.; Tanaka, T.; Kon, H.; Oota, K.; Kawaguchi, A.; Maki, Y.; Sato, T.
 J. Dairy Sci. 76, 946-955, 1993
 A/Title: Separation and characterization of the C-terminal half molecule of bovine lacto
 A/Reference number: A56659; MUID:93253156; PMID:8486845
 A/Accession: A56659
 A/Molecule type: protein
 A/Residues: 20-25;302-308;359-366, 'X', 368-376, 'X', 378 <SHI>
 C/Superfamily: transferrin; transferrin repeat homology
 C/Keywords: duplication; glycoprotein; iron; iron binding; metalloprotein; milk
 F/1-19/Domin: signal sequence #status predicted <SIG>
 F/20-708/Domain: lactoferrin repeat homology <MAT>
 F/20-355/Domain: transferrin repeat homology <TRH>
 F/356-60/Region: antimicrobial
 F/359-696/Domain: transferrin repeat homology <TRH2>
 F/28-64;134-217;176-192;179-200;189-202;250-264;367-399;377-390;424-703;444-666;476-551,
 F/38-55/Diulfide bonds: #status predicted
 F/79,111,211,277/binding site: iron (Asp, Tyr, Tyr, His) #status experimental
 F/140/binding site: carbonate (Arg) #status experimental
 F/252,300,367,495,564/binding site: carbohydrate (Asn) (covalent) #status predicted
 F/141,452,545,614/binding site: iron (Asp, Tyr, Tyr, His) #status experimental
 F/482/binding site: carbonate (Arg) #status experimental

Query Match

74.4%; Score 122; DB 1; Length 708;

Best Local Similarity 72.4%; Pred. No. 7.4e-10;
 Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 1 PENSCTQWQPRNRKTCAPSITCIRRTSA 29
 DB 33 PEFKCRRWQWRMKKIGAPSTICVRRAFA 61
 RESULT 3
 TPRUL
 lactoferrin precursor [validated] - human
 M/Alternate names: lactoferrin
 C/Species: Homo sapiens (man)
 C/Date: 31-Mar-1992 #sequence revision 21-Nov-1997 #text change 08-Dec-2000
 C/Accession: G01394; S11228; A45401; S10324; S15853; S20841; S07160; A61169; A31000; S74
 R/Cho, Y.
 submitted to the EMBL Data Library, March 1994
 A/Reference number: G06820
 A/Accession: G01394
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-711 <CHO>
 A/Cross-references: EMBL:U07643; NID:9467236; PIDN:AAB60324.1; PID:9467237
 R/Rey, M.W.; Woloshuk, S.L.; deBoer, H.A.; Pieper, F.R.
 Nucleic Acids Res. 18, 5288, 1990
 A/Title: Complete nucleotide sequence of human mammary gland lactoferrin.
 A/Reference number: S11228; MUID:90384839; PMID:2402455
 A/Accession: S11228
 A/Molecule type: mRNA
 A/Residues: 1-148, 'T', 150-422, 'C', 424-711 <REX>
 A/Cross-references: EMBL:X53961; NID:934415; PIDN:CAA37914.1; PID:934416
 R/Teng, C.T.; Liu, Y.; Yang, N.; Walmer, D.; Panella, T.
 Mol. Endocrinol. 6, 1969-1981, 1992
 A/Title: Differential molecular mechanism of the estrogen action that regulates lactofer
 A/Reference number: A45401; MUID:93125571; PMID:1480183
 A/Accession: A45401
 A/Molecule type: DNA
 A/Residues: 1-15 <TEN>
 A/Cross-references: GB:S52659; NID:9263311; PIDN:AAB24877.1; PID:9263312
 A/Experimental source: placenta
 A/Note: sequence extracted from NCBI backbone (NCBI:122202)
 R/Powell, M.J.; Ogden, J.B.
 Nucleic Acids Res. 18, 4013, 1990
 A/Title: Nucleotide sequence of human lactoferrin cDNA.
 A/Reference number: S10324; MUID:90326549; PMID:2374734
 A/Accession: S10324
 A/Molecule type: mRNA
 A/Residues: 3-711 <POW>
 A/Cross-references: EMBL:X52941; NID:934411; PIDN:CAA37116.1; PID:934412
 R/Stowell, K.M.; Rado, T.A.; Funk, W.D.; Tweedie, J.W.
 Biochem. J. 276, 349-355, 1991
 A/Title: Expression of cloned human lactoferrin in baby-hamster kidney cells.
 A/Reference number: S15853; MUID:91264786; PMID:2045066
 A/Accession: S15853
 A/Status: nucleic acid sequence not shown; not compared with conceptual translation
 A/Molecule type: mRNA
 A/Residues: 20-31 <STI>
 A/Accession: S20841
 A/Molecule type: protein
 A/Residues: 20-28, 'X', 30-31 <ST2>
 R/Rado, T.A.; Wei, X.; Benz Jr., E.J.
 Blood 70, 989-993, 1987
 A/Title: Isolation of lactoferrin cDNA from a human myeloid library and expression of mR
 A/Reference number: S07160; MUID:88001031; PMID:3477300
 A/Accession: S07160
 A/Molecule type: mRNA
 A/Residues: 435-487, 'A', 489-711 <RAD>
 A/Cross-references: EMBL:M18642; NID:9186815; PIDN:AAA66655.1; PID:9386855
 R/Panella, T.C.; Liu, Y.; Huang, A.T.; Teng, C.T.
 Cancer Res. 51, 3037-3043, 1991
 A/Title: Polymorphism and altered methylation of the lactoferrin gene in normal leukocyte
 A/Reference number: A61169; MUID:91235214; PMID:1674448
 A/Accession: A61169

A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 32701,'SMKPVN' <PAN>
A:Experimental source: normal breast tissue
R:Meiz-Soutigue, M.H.; Jolles, U.; Mazurier, J.; Schoentgen, F.; Legrand, D.; Spik, G.,
bur. J. Biochem. 145, 659-666, 1984
A>Title: Human lactoferrin: amino acid sequence and structural comparisons with oth
A:Reference number: A11000; MUID:85076667; PMID:6510420
A:Accession: A31000
A:Molecule type: protein
A:Residues: 20-140,141-159,171-203,'L',205,'K',207-208,'K',210-385,'Q',387-391,'W',393-4
A:Note: this is the final paper in a series
R:Henou, G.; Hoegdall, B.V.; Barkholt, V.; Nørskov, L.
bur. J. Biochem. 241, 303-308, 1996
A>Title: Lactoferrin: similarity to diamine oxidase and purification by aminohexyl affini
A:Reference number: S74119; MUID:97054624; PMID:8898921
A:Accession: S74119
A:Molecule type: protein
A:Residues: 'G',23-24,'R',26-27,'XX',30-32 <HOU>
A:Experimental source: neutrophil granulocytes
C:Genetics:
A:Gene: GDB:LTF
A:Cross-references: GDB:119368; OMIM:150210
A:Map position: 3q21-3q23
C:Superfamily: transferrin; transferrin repeat homology
C:Keywords: duplication; glycoprotein; iron binding; milk
F:1-19/Domain: signal sequence #status predicted <SIG>
F:21-711/Product: lactotransferrin #status experimental <MAT>
F:21-356/Domain: transferrin repeat homology <TRH>
F:360-699/Domain: transferrin repeat homology <TRH>
F:129-65,39-56,135-218,117-193,190-201,251-265,503-597,595-609/Disulfide bonds: #status
F:157,488/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:368-400,378-391,425-706,447-669,479-554,513-527,524-537,647-652/Disulfide bonds: #stat

Query Match
Best Local Similarity 51.0%; Score 100; DB 1; Length 711;
Matches 16; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

DB 1 PMSKCYOWQRMKLGAPSTICRTS 28
34 PEATKCFQWQRMKRVGPVSCIKRDS 61

RESULT 4
SS2107
lactoferrin - sheep (fragment)
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 14-Jul-1995 #sequence_revision 05-Dec-1997 #text_change 05-Dec-1997
C:Accession: SS2107
R:Qian, Z.Y.; Jolles, P.; Migliore-Samouri, D.; Fiat, A.M.
Biochim. Biophys. Acta 1243, 25-32, 1995
A>Title: Isolation and characterization of sheep lactoferrin, an inhibitor of platelet a
A:Reference number: SS2107; MUID:95127729; PMID:7827104
A:Accession: SS2107
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-33 <QIA>
C:Superfamily: transferrin; transferrin repeat homology
C:Keywords: duplication

Query Match
Best Local Similarity 55.5%; Score 91; DB 2; Length 33;
Matches 15; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DB 1 PMSKCYOWQRMKLGAP 19
14 PEGSKCYOWQKMKRKLGRP 32

RESULT 5
128438
lactoferrin precursor - mouse
N:Alternate names: lactotransferrin

C:Species: Mus musculus (house mouse)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C/Accession: A28438; A41205
R:Pentecost, B.T.; Teng, C.T.
J: Biol. Chem. 262, 10134-10139, 1987
A>Title: Lactotransferrin is the major estrogen inducible protein of mouse uterine secr.
A:Reference number: A92596; PMID:87280033; PMID:3611056
A:Accession: A28438
A:Molecule type: mRNA
A:Residues: 3707 <EN>
A/Cross-references: EMBL:J03298
R:Liu, Y.; Teng, C.T.
J: Biol. Chem. 266, 21880-21885, 1991
A>Title: Characterization of estrogen-responsive mouse lactoferrin promoter.
A:Reference number: A41205; PMID:92042099; PMID:1939212
A:Accession: A41205
A:Molecule type: DNA
A:Residues: 1-15 <LU>
A/Cross-references: GB:M74778
C/Superfamily: transferrin; transferrin repeat homology
C/Keywords: duplication; glycoprotein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-70/Product: lactotransferrin #status predicted <MAT>
F:358-695/Domain: transferrin repeat homology <TRH>
F:494/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 51.2%; Score 84; DB 1; Length 707;
Best Local Similarity 44.4%; Freq. No. 0.00027;
Matches 12; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 EMSKCYQMRRKKLGAPSIPTCIIRTS 28
Db 33 EEKCLRWQNEKRKYGSPPLCVKKSS 59
| | | | | | | | : : : :
| | | | | | | | : : : :

RESULT 6
A43543
lactoferrin precursor - pig
C/Species: Sus scrofa domestica (domestic pig)
C/Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 04-Mar-2000
C/Accession: A45543; S24173
R:Alexander, L.U.; Levine, W.B.; Teng, C.T.; Beattie, C.W.
Ann. Genet. 23, 251-256, 1992
A>Title: Cloning and sequencing of the porcine lactoferrin cDNA.
A:Reference number: A45543; PMID:92367939; PMID:1503259
A:Accession: A45543
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-703 <LR>
A:Experimental source: mammary gland
R:Pydon, D.P.; O'Malley, B.R.; Saucedo, O.; Lee, T.; Heaton, D.R.; Conneely, O.M.
Biochim. Biophys. Acta 1132, 97-99, 1992
A>Title: Nucleotide and primary amino acid sequence of porcine lactoferrin.
A:Reference number: S24173; PMID:92379101; PMID:1511016
A:Accession: S24173
A:Molecule type: mRNA
A:Residues: 1-11, 'W', '13-50', 'I', '52-84', 'G', '86-120', 'L', '121-130', 'I', '132-282', 'S', '284-571', 'Q'
A/Cross-references: EMBL:M92089; NID:g164613; PIDN:AAA31102.1; PID:g164614
A:Experimental source: mammary gland
C/Superfamily: transferrin; transferrin repeat homology
C/Keywords: duplication; glycoprotein; iron; iron binding; metalloprotein; milk
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-70/Product: lactoferrin #status predicted <MAT>
F:20-350/Domain: transferrin repeat homology <TRH>
F:36-48/Region: antimicrobial
F:354-691/Domain: transferrin repeat homology <TRH2>
F:28-62, 38-53, 129-212, 171-187, 184-195, 245-259, 362-394, 377-385, 419-698, 439-661, 471-546, 454-477, 107, 206, 267/Binding site: iron (asp, tyr, tyr, his) #status predicted
F:135/Binding site: carbonate (arg) #status predicted
F:409, 447, 540, 609/Binding site: iron (asp, tyr, tyr, his) #status predicted
F:477/Binding site: carbonate (arg) #status predicted
F:490/Binding site: carbohydrate (asn) (covalent) #status predicted

```

Query Match 40.2%; Score 66; DB 2; Length 703;
Best Local Similarity 48.1%; Pred. No. 0.13;
Matches 13; Conservative 5; Mismatches 7; Indels 2; Gaps 1;

OY 2 EWSKCYQWQRMRKLGAPSTICIRRTS 28
:|||||:|:|:|
Db 34 EYKSCROWQMSKIRRTN--PMFCIRRAS 58

RESULT 7
147228
C:carbonic anhydrase II inhibitor (transferrin homolog) precursor - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 20-Aug-1999
C:Accession: I47228
R:Roush, E.D.; Flierke, C.A.
Biochemistry 31, 12536-12542, 1992
A>Title: Purification and characterization of a carbonic anhydrase II inhibitor from porcine erythrocytes
A:Reference number: I47228; PMID:1463741
A:Accession: I47228
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-704 <ROU>
A:Cross-references: EMBL:U36916; NID:g1016329; PIDD:AAB58956.1; PID:g1016330
C:Genetics:
A:Gene: PICA
C:Superfamily: transferrin; transferrin repeat homology
C:Keywords: duplication
F:20-350/Domain: transferrin repeat homology <TRH1>

Query Match 35.7%; Score 58.5; DB 2; Length 704;
Best Local Similarity 36.7%; Pred. No. 1.5;
Matches 11; Conservative 8; Mismatches 8; Indels 3; Gaps 1;

OY 2 EWSKCYQWQRMRKLGAPSTICIRRTS 28
:|||||:|:|:|
Db 34 EASKCSFRHNMKILPVEGPHVSCVKTSS 63

RESULT 8
EBS5813
Hypotheetical protein Z3045 [imported] - Escherichia coli (strain O157:H7, substrain EDL93)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Dec-2001
C:Accession: EBS5813
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: EBS5813
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-271 <STO>
A:Cross-references: GI:AB005174; NID:g12516061; PIDD:AAG56969.1; GSPDB:GN00145; UWGP:Z3045
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3045
C:Superfamily: Pyrococcus horikoshii hypotheetical protein PH0926

Query Match 35.4%; Score 58; DB 2; Length 271;
Best Local Similarity 35.7%; Pred. No. 0.71;
Matches 10; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

OY 2 EWSKCYQWQRMRKLGAPSTICIRRTSA 29
:|||||:|:|:|
Db 25 DWQPAAPWLSRLREANVPIILCSKTS 52

Hypotheetical protein EBS2693 [imported] - Escherichia coli (strain O157:H7, substrain EDL93)

```

C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 14-Dec-2001
C:Accession: E90965
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
G:Gotoh, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E90965
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-271 <HAV>
A:Cross-references: GB:AB000007; PIDN:BAH36116; PID:G13362161; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509592
C:Gene: ECA2693
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0926

Query Match 35.4%; Score 58; DB 2; Length 271;
Best Local Similarity 35.7%; Pred. No. 0.71;
Matches 10; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

RESULT 10
G64959
hypothetical protein b1955 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: G64959
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A:Rose, D.U.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G64959
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-271 <BLAT>
A:Cross-references: GB:AB000287; GB:U00096; MID:G1788257; PIDN:AACT5021.1; PID:G1788265,
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0926

Query Match 35.4%; Score 58; DB 2; Length 271;
Best Local Similarity 35.7%; Pred. No. 0.71;
Matches 10; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

RESULT 11
AB0754
conserved hypothetical protein STY2193 [imported] - Salmonella enterica subsp. enterica
C:Species: Salmonella enterica subsp. enterica serovar Typh
C:Note: This species has also been called Salmonella typh
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AB0754
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moutle, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AB0754
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-271 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05733.1; PID:G16503226; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY2193
 C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0926

Query Match 35.4%; Score 58; DB 2; Length 271;
 Best Local Similarity 35.7%; Pred. No. 0.71;
 Matches 10; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

DB 25 EWQPAAPMLTRHESGVPLCSKTA 52

RESULT 12
 TFHM
 melanotransferrin precursor - human
 N:Alternate names: melanoma-associated antigen gp95/p97
 C:Species: Homo sapiens (man)
 C>Date: 31-Mar-1998 #sequence_revision 31-Mar-1998 #text_change 28-Jan-2000
 C:Accession: A23814; A60925
 R:Rose, T.M.; Plowman, G.D.; Teplow, D.B.; Dreyer, W.J.; Hellstrom, K.E.; Brown, J.P.
 Proc. Natl. Acad. Sci. U.S.A. 83, 1261-1265, 1986
 A:Title: Primary structure of the human melanoma-associated antigen p97 (melanotransferrin)
 A:Reference number: A23814; MUID:8619285; PMID:2419904
 A:Accession: A23814
 A:Molecule type: mRNA
 A:Residues: 1-738 <ROS>
 A:Cross-references: EMBL:M12154; NID:G189515; PIDN:AA59992.1; PID:G189518
 A:Experimental source: melanoma
 R:Funkhwa, K.S.; Funkhwa, K.; Real, F.X.; Old, L.T.; Lloyd, K.O.
 J. Exp. Med. 169, 585-590, 1989
 A:Title: A unique antigenic epitope of human melanoma is carried on the common melanoma
 A:Reference number: A60925; MUID:8904252; PMID:2463331
 A:Accession: A60925
 A:Molecule type: protein
 A:Residues: 20-25; 'X', 27-28; 'X', 30 <FUR>
 C:Comment: This protein is found predominantly in human melanomas and in certain fetal
 C:Genetics:
 A:Gene: GDB:MFT2
 A:Cross-references: GDB:119387; OMIM:155750
 A:Map position: 3q28-3q29
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keywords: blocked carboxyl end; duplication; glycoprotein; lipoprotein; membrane probe
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:18-360/Domain: transferrin repeat homology <TRH1>
 F:20-710/Product: melanotransferrin #status predicted <WTF>
 F:361-709/Domain: transferrin repeat homology <TRH2>
 F:711-738/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F:38, 135, 515/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:710/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature form)

Query Match 35.1%; Score 57.5; DB 1; Length 738;
 Best Local Similarity 43.3%; Pred. No. 2.2;
 Matches 13; Conservative 3; Mismatches 13; Indels 1; Gaps 1;

DB 31 PEQKCGNMBARFAGIOPSLCVRGTA 60

RESULT 13
 A47353
 phosphatidylethanolamine N-methyltransferase (EC 2.1.1.17), 20K - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 08-May-1995 #sequence_revision 12-May-1995 #text_change 05-May-2000
 C:Accession: A47353
 R:Gul, Z.; Vance, J.E.; Chen, M.H.; Voelker, D.R.; Vance, D.E.
 J. Biol. Chem. 268, 1655-1666, 1993
 A:Title: Cloning and expression of a novel phosphatidylethanolamine N-methyltransferase.
 A:Reference number: A47353; MUID:93346418; PMID:8344945
 A:Accession: A47353

A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-199 <CUI>
 A:Cross-references: GB:U14441; NID:G310194; PIDN:AA03154.1; PID:G310195
 C:Superfamily: phosphatidyl-N-methylethanolamine N-methyltransferase
 C:Keywords: membrane protein; methyltransferase; S-adenosylmethionine

Query Match 31.1%; Score 51; DB 2; Length 199;
 Best Local Similarity 36.0%; Pred. No. 5.6;
 Matches 9; Conservative 6; Mismatches 6; Indels 4; Gaps 1;

DB 29 WNVVAREWQRTRLSPAFSPYLAC 53

RESULT 14
 C70655
 probable monooxygenase - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: C70655
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
 ; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: C70655
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-489 <COL>
 A:Cross-references: GB:283664; GB:AL123456; NID:G3261687; PIDN:CA806212.1; PID:G31250;
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: RV3854c

Query Match 31.1%; Score 51; DB 2; Length 489;
 Best Local Similarity 57.1%; Pred. No. 13;
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DB 3 WSKCYQMRMRKL 16

RESULT 15
 S33761
 transferrin precursor - horse
 N:Alternate names: growth-promoting factor
 C:Species: Equus caballus (domestic horse)
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
 R:Carpenster, M.A.; Broad, T.R.
 Biochim. Biophys. Acta 1173, 230-232, 1993
 A:Title: The cDNA sequence of horse transferrin.
 A:Reference number: S33761; MUID:93277958; PMID:8504171
 A:Accession: S33761
 A:Molecule type: mRNA
 A:Residues: 1-706 <CAR>
 A:Cross-references: EMBL:M69020; NID:G164242; PIDN:AA30958.1; PID:G164243
 A:Experimental source: liver; developmental stage adult
 R:Yoshinari, K.; Yuasa, K.; Iga, F.; Matsuura, A.
 Biochim. Biophys. Acta 1010, 38-34, 1989
 A:Title: A growth-promoting factor for human myeloid leukemia cells from horse serum (cf
 A:Reference number: S02145; MUID:89076897; PMID:2902248
 A:Accession: S02145
 A:Molecule type: protein
 A:Residues: 20-35; 'X', 37; 'X', 39-40; 'X', 43-44 <YOS>
 C:Complex: monomer
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keywords: duplication; glycoprotein; iron transport; metal binding; plasma

F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-706/Product: transferrin #status experimental <MAT>
F:358-694/Domain: transferrin repeat homology <TRH2>
F:26-64,36-55,134-215,174-190,177-198,187-200,248-262,360-623,366-398,376-389,423-701,44
Query Match 30.5%; Score 50; DB 2; Length 706;
Best Local Similarity 34.5%; Pred. No. 26;
Matches 10; Conservative 7; Mismatches 10; Indels 2; Gaps 1;
QY 2 EMSKCYQWQRMRL--GAPSTCIARTS 28
| | | | | : : : : : |
Db 32 EVSKCASFDSMSKSIVPAPPLVACVKRTS 60

Search completed: July 30, 2003, 16:31:17
Job time : 20.4702 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:09:50 ; Search time 57.1369 Seconds
(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-5

Perfect score: 164
Sequence: 1 PEMSKCYQWQRMKLGAPSTICIRRTSA 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	164	100.0	29	21	AAV88341
2	164	100.0	708	20	AAAG80774
3	163	99.4	708	20	AAAG80775
4	136	82.9	25	21	AAV883112
5	136	82.9	25	21	AAV883013
6	122	74.4	689	21	AAV49270
7	122	74.4	708	17	AAW09343
8	122	74.4	708	19	AAW57318
9	122	74.4	708	20	AAW86022

10	122	74.4	708	21	AAV88341	Amino acid sequenc
11	122	74.4	708	22	AAE02342	Chronic hepatitis
12	122	74.4	708	22	AAE02342	Bovine lactoferrin
13	120	73.2	32	15	AAV44845	Lactoferrin-relate
14	120	73.2	32	15	AAV44845	Lactoferrin-relate
15	120	73.2	32	15	AAV57466	Lactoferrin-relate
16	120	73.2	32	16	AAV84703	Bovine lactoferrin
17	120	73.2	32	17	AAV80268	Anti-parasitic lac
18	120	73.2	32	17	AAV91856	Lactoferrin-relate
19	120	73.2	32	17	AAV90610	Lactoferrin-relate
20	120	73.2	32	17	AAV87625	Lactoferrin-relate
21	120	73.2	32	18	AAV26154	Lactoferrin-relate
22	120	73.2	32	18	AAV14040	Anti-parasitic pep
23	120	73.2	32	19	AAV70314	Thrombus formation
24	120	73.2	32	19	AAV53229	Lactoferrin hydrol
25	120	73.2	32	21	AAV56543	Apoptosis inducer
26	120	73.2	32	21	AAV56543	Peptide SEQ ID NO:
27	120	73.2	38	14	AAV44844	Lactoferrin-relate
28	120	73.2	38	15	AAV44844	Lactoferrin-relate
29	120	73.2	38	15	AAV57465	Lactoferrin-relate
30	120	73.2	38	16	AAV84702	Bovine lactoferrin
31	120	73.2	38	16	AAV80267	Anti-parasitic lac
32	120	73.2	38	17	AAV91855	Lactoferrin-relate
33	120	73.2	38	17	AAV90609	Lactoferrin-relate
34	120	73.2	38	17	AAV87624	Lactoferrin-relate
35	120	73.2	38	18	AAV26153	Lactoferrin-relate
36	120	73.2	38	18	AAV14039	Anti-parasitic pep
37	120	73.2	38	19	AAV70313	Thrombus formation
38	120	73.2	38	19	AAV53228	Lactoferrin hydrol
39	120	73.2	38	19	AAV44845	Apoptosis inducer
40	120	73.2	38	21	AAV56542	Peptide SEQ ID NO:
41	115	70.1	704	22	AAV51535	Recombinant lactof
42	105	64.0	51	17	AAV91194	Lactoferrin decomp
43	105	64.0	52	16	AAV87909	Bovine lactoferrin
44	100	61.0	36	14	AAV44846	Lactoferrin-relate
45	100	61.0	36	17	AAV10517	Lactoferrin-relate

ALIGNMENTS

RESULT 1	AAV88341	standard, peptide, 29 AA.
ID	AAV88341;	
AC	AAV88341;	
XX		
XX		
DT	14-JUL-2000 (first entry)	
DE	Peptide derived from an internal fragment of goat whey.	
XX		
XX		
KW	Peptide production; biological fluid; milk; whey; blood;	
KW	antibacterial peptide; lactoferrin; antiviral; antitumour activity.	
OS	Capra sp.	
XX		
XX		
FT	Key	Location/Qualifiers
FT	Disulfide-bond	6..23 /note="Optionally not present"
XX		
XX		
XX		
XX		
XX		
PD	23-MAR-2000.	
XX		
XX		
PF	15-SEP-1999;	99WC-EP07002.
XX		
PR	15-SEP-1998;	98EP-0203107.
PR	08-JUN-1999;	99EP-0201615.
XX		
XX		
PA	(NIZO-) NIZO FOOD RES.	
XX		
XX		
PI	Visser S, Recio I;	
XX		

DR WPI; 2000-271377/23.

XX Novel process for producing peptides with e.g. antimicrobial activity
PT from biological fluids such as milk, whey or blood comprises contacting
PT fluid with chromatographic medium to adsorb peptide domain of interest
XX

PS Claim 14; Page 22; 41pp; English.

CC This sequence represents a peptide derived from an internal fragment of
CC goat whey. The peptide is an example of a peptide with antibacterial
CC activity that can be produced by the process of the invention. The
CC invention relates to a process for producing peptides from biological
CC fluids. The process comprises chromatography of the biological fluid, in
CC situ hydrolysis of selectively bound peptides, washing to remove unbound
CC peptide, and elution of the peptides of interest. The process is used for
CC producing peptides from biological fluids, such as milk, whey or blood.
CC For example, the process can be used to produce antibacterial peptides
CC derived from lactoferrin, using cheese whey as a starting material. The
CC peptides obtained have preferably antimicrobial and/or antiviral and/or
CC antitumour activity. The process of the invention is relatively simple
CC and generally economically and technically more attractive than those
CC methods previously used. The method provides high yield peptides with a
CC selected activity of interest without the need for intermediate
CC purification of the precursor protein.

XX Sequence 29 AA;

Query Match 100.0%; Score 164; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 8.4e-17;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMSKCYQOMRRMKLGAPSTICIRRTSA 29
DB 1 PMSKCYQOMRRMKLGAPSTICIRRTSA 29

RESULT 2

AAG80774 AAG80774 standard; Protein; 708 AA.

AC AAG80774;

DT 19-APR-2002 (first entry)

DE Goat lactoferrin-associated protein #1.

XX Primer; lactoferrin; antibiotic; goat.

OS Capra sp.

XX Key Location/Qualifiers

FT Misc-difference 360 /note= "Encoded by CGG"

XX KR98043944-A.

XX 05-SEP-1998.

XX 05-DEC-1996; 96KR-0061925.

XX 05-DEC-1996; 96KR-0061925.

XX (KOAD) KOREA ADV INST SCI & TECHNOLOGY.

XX Yoo DY, Nam MS, Lee TH, Yoo SR, Kim SJ, Lee GG;

XX WPI; 1999-465631/39.

XX N-PSDB; ABA97187.

XX Antibiotic lactoferrin derived from Korean black goat and gene thereof -
XX NoAbstract

PS Disclosure; Page 6-9; 11pp; Korean.

XX This sequence represents the goat antibiotic lactoferrin described in the
CC disclosure of the invention.
XX

XX Sequence 708 AA;

Query Match 100.0%; Score 164; DB 20; Length 708;
Best Local Similarity 100.0%; Pred. No. 2.6e-15;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMSKCYQOMRRMKLGAPSTICIRRTSA 29
DB 33 PMSKCYQOMRRMKLGAPSTICIRRTSA 61

RESULT 3

AAG80775 AAG80775 standard; Protein; 708 AA.

AC AAG80775;

DT 19-APR-2002 (first entry)

DE Goat lactoferrin-associated protein #2.

XX Primer; lactoferrin; antibiotic; goat.

OS Capra sp.

XX Key Location/Qualifiers

FT Misc-difference 360 /note= "Encoded by CGG"

XX KR98043944-A.

XX 05-SEP-1998.

XX 05-DEC-1996; 96KR-0061925.

XX 05-DEC-1996; 96KR-0061925.

XX (KOAD) KOREA ADV INST SCI & TECHNOLOGY.

XX Yoo DY, Nam MS, Lee TH, Yoo SR, Kim SJ, Lee GG;

XX WPI; 1999-465631/39.

XX NoAbstract

XX Antibiotic lactoferrin derived from Korean black goat and gene thereof -

XX Disclosure; Page 6-9; 11pp; Korean.

XX This sequence represents the goat antibiotic lactoferrin described in the

XX disclosure of the invention.

XX Sequence 708 AA;

Query Match 99.4%; Score 163; DB 20; Length 708;
Best Local Similarity 96.6%; Pred. No. 3.6e-15;
Matches 28; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMSKCYQOMRRMKLGAPSTICIRRTSA 29
DB 33 PMSKCYQOMRRMKLGAPSTICIRRTSA 61

RESULT 4

AA183112 AA183112 standard; peptide; 25 AA.

XX AAY83112;
XX AAY83112;

```

DT 24-JUN-2000 (first entry)
XX Synthetic lactoferrin fragment.
DE Lactoferrin; antimicrobial; cytostatic; infection; tumour;
XX sterilisation; wound healing; spermicide; goat.
XX Synthetic.
OS Capra hircus.
XX WO200012541-A2.
XX
XX 09-MAR-2000.
XX
XX 31-AUG-1999; 99WO-GB02850.
XX
XX 28-AUG-1998; 98GB-0018938.
XX
XX (ALPH-) ALPHARMA AS.
XX (GARD/) GARDNER R.
XX
XX Svendsen JS, Rekdal O, Sveinbjornson B, Vorland L;
XX MPI, 2000-270793/23.
XX
XX Cytotoxic peptides useful as medicament for treating tumors and
XX bacterial infections, comprises one or more non-genetic bulky and
XX lipophilic amino acids
XX
XX Example 1; Figure 1; 114pp; English.
XX
XX Cytotoxic 7-25 mer lactoferrin peptides which comprise three or more
XX cationic residues and have one or more non-genetic bulky and
XX lipophilic amino acids have cytostatic and antimicrobial activity.
XX The peptides are useful as medicament for treating bacterial
XX infections and tumors. They are also useful for sterilising
XX agents for materials susceptible to microbial contamination. They may
XX also be used as promoters of wound healing and spermicides. The small
XX size of the peptides gives them an increased half life and allows
XX efficient biodelivery. The peptides can be administered without need
XX for an injection, such as by inhalation or by absorption across the
XX blood capillaries of the nasal passages. This synthetic lactoferrin
XX peptide corresponds to amino acids 17-41 of caprine lactoferrin.
XX
XX Sequence 25 AA;
XX
XX Query Match 82.9%; Score 136; DB 21; Length 25;
XX Best Local Similarity 96.0%; Pred. No. 8, 8e-13;
XX Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 4 SKCYQWRMRKLGAPSTICIRRTS 28
XX |||||
XX 1 SKCYQWRMRKLGAPSTICIRRTS 25
XX
XX RESULT 5
XX AAY85013
XX ID AAY85013 standard; peptide; 25 AA.
XX
XX AAY85013;
XX
XX 20-JUN-2000 (first entry)
XX
XX Caprine lactoferrin peptide LFC(17-41) amino acid sequence.
XX
XX Lactoferrin; lactoferricin; antibacterial; bacterial infection;
XX bacterial growth; solid tumour; treatment; goat.
XX
XX Capra sp.
XX
XX WO200012542-A2.
XX
XX 09-MAR-2000.
XX

```

```

XX 31-AUG-1999; 99WO-GB02851.
XX
XX 28-AUG-1998; 98GB-0018938.
XX
XX (ALPH-) ALPHARMA AS.
XX (GARD/) GARDNER R.
XX
XX Svendsen JS, Rekdal O, Sveinbjornson B, Vorland L;
XX MPI, 2000-256582/22.
XX
XX Modified cytotoxic lactoferrin peptide useful for preparing
XX pharmaceutical composition against bacterial infections and tumors and
XX for inhibiting bacterial growth
XX
XX Example 1; Fig 1; 112pp; English.
XX
XX This sequence represents a caprine lactoferricin peptide. Lactoferrin is
XX a weak iron transporter which also shows weak antibacterial effect.
XX Lactoferricin B (LFB) is a peptide produced when bovine lactoferrin is
XX digested with pepsin. LFB is more active than lactoferrin. The two
XX tryptophan residues at positions 6 and 8 in LFB 17-31 cannot be
XX substituted with alanine without loss of antibacterial activity. The
XX invention relates to a modified cytotoxic lactoferrin peptide of 7-25
XX amino acids in length with three cationic residues and one or more extra
XX bulky and lipophilic amino acids, than its native peptide. The invention
XX also relates to the esters, amides, salts and cyclic derivatives of the
XX modified peptide. The modified lactoferrin peptides of the invention are
XX useful for treating bacterial infections and tumors, for preparing
XX medicaments and for inhibiting bacterial growth. LFB and its fragments
XX are useful for manufacturing medicaments for treating solid tumors. The
XX modified peptides have a very small size, which is suitable for
XX biodelivery. Due to the smaller size, the circulating half-life of the
XX peptide is increased as they are less vulnerable to endopeptidases. The
XX observed cytolytic effect of the modified lactoferrin peptides in tumors
XX is not species specific and they have greater utility in treating human
XX tumors.
XX
XX Sequence 25 AA;
XX
XX Query Match 82.9%; Score 136; DB 21; Length 25;
XX Best Local Similarity 96.0%; Pred. No. 8, 8e-13;
XX Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 4 SKCYQWRMRKLGAPSTICIRRTS 28
XX |||||
XX 1 SKCYQWRMRKLGAPSTICIRRTS 25
XX
XX RESULT 6
XX AAY49270
XX ID AAY49270 standard; protein; 689 AA.
XX
XX AAY49270;
XX
XX 07-FEB-2000 (first entry)
XX
XX Lactoferrin sequence.
XX
XX Adhesion inhibition; enteropathogenic; Escherichia coli; lactoferrin;
XX KW medicament; drink; feed; food poisoning.
XX
XX Undenitrified.
XX
XX JP11292789-A.
XX
XX 26-OCT-1999.
XX
XX 03-APR-1998; 98JP-0107167.
XX
XX 03-APR-1998; 98JP-0107167.
XX

```


XX	Lactoferrin; antibiotic; chelate; mastitis; bowel disorder;	
KW	disease; bacteria; yeast; fungi; disinfection; drug; foodstuff;	
KW	cosmetic; toiletries.	
XX		
OS	Unidentified.	
XX		
Key	Location/Qualifiers	
PT	Disulfide-bond 10..27	
XX		
PN	NO93J14640-A1.	
XX		
PD	05-AUG-1993.	
XX		
PP	30-NOV-1992; 92WO-JP01563.	
XX		
PR	23-JAN-1992; 92JP-0032660.	
PR	11-MAR-1992; 92JP-0052943.	
PR	30-SEP-1992; 92JP-0262143.	
PR	30-SEP-1992; 92JP-0262559.	
XX		
PA	(MORG) MORINAGA MILK IND CO LTD.	
PI	Belamy W, Fukuwatari Y, Kawase K, Shimamura S;	
PI	Takase M, Tokitay, Tomita M, Wakabayashi H, Yamuchi K;	
XX	WPI; 1993-258265/32.	
DR		
XX		
PT	Antibacterial agent comprising decomposition products of	
PT	lactoferrin - with chelate e.g. EDTA alcohol and/or antibiotic	
PT	e.g. penicillin, also useful against yeast and fungi	
XX		
PS	Disclosure; Page 87; 100pp; Japanese.	
XX		
CC	Lactoferrin-related peptides are used in new antibacterial compn.	
CC	The compn. is highly effective against a broad range of bacteria,	
CC	yeasts and fungi. It can be used therapeutically (internal and	
CC	external application), e.g. for mastitis, bowel disorders, urinary	
CC	infections, etc. It can also be used for the disinfection and	
CC	protection of drugs, foodstuffs, cosmetics and toiletries and	
CC	household items (such as kitchen towels and toilet paper).	
CC	(Updated on 09-JAN-2003 to add missing OS field.)	
CC	(Updated on 25-MAR-2003 to correct PN field.)	
XX		
XX	Sequence 32 AA;	
XX		
Query Match	73.2%; Score 120; DB 14; Length 32;	
Best Local Similarity	76.9%; Pred. NO. 2.5e-10;	
Matches 20; Conservative	3; Mismatches 3; Indels 0; Gaps 0;	
0Y	1 PMSKRYQWQRRKKLGASSTCTCR 26	
	: : :	
DB	5 PSMFKCRFQWRRKKLGASSTCYVR 30	
XX		
RESULT 14		
ID	AAR48534 standard; peptide; 32 AA.	
XX	AAR48534:	
AC		
XX		
DT	25-MAR-2003 (updated)	
DT	10-AUG-1994 (first entry)	
XX		
XX		
DB	Lactoferrin derived peptide #28.	
XX		
KW	Decomposition; lactoferrin; digestion; enzyme; pepsin; trypsin;	
KW	antioxidant; oxidation; inhibitor; vitamin E; ascorbic acid;	
KW	vitamin A; beta-carotene; superoxide dismutase; coenzyme Q;	
KW	lipid oxidation; foodstuff; drugs; health food; toiletries; cosmetics.	
XX		
XX	Bos taurus.	
XX		

XX	MO9403555-A1.
XX	PD 17-FEB-1994.
XX	PF 04-AUG-1993; 93WO-JP01090.
XX	PR 07-AUG-1992; 92JP-0211335.
XX	(MORG) MORINAGA MILK IND CO LTD.
XX	
PI	Bellamy WR, Fukuwatari Y, Kawase K, Shimamura S,
PI	Takase M, Tokiday, Tomita W, Wakabayashi H, Yamauchi K;
DR	WPI; 1994-065650/08.
XX	
PT	Antioxidant peptide lactoferrin decomposition product - prevents
PT	oxidation of lipid(s) in foodstuffs and drugs without affecting
PT	their taste
PS	
CS	Claim 3; Page 39; 47pp; Japanese.
CC	The sequences given in AAR48507-37 are peptides derived by the
CC	decomposition of lactoferrin, pref. by digestion with an enzyme, eg.
CC	pepsin or trypsin. These peptides may be used in an antioxidant
CC	composition which may also contain an oxidation inhibitor such as
CC	vitamin E, ascorbic acid, vitamin A, beta-carotene, superoxide
CC	dismutase or coenzyme Q. The antioxidant prevents lipid oxidation
CC	in foodstuffs, drugs, health foods, toiletries and cosmetics.
CC	(Updated on 25-MAR-2003 to correct PN field.)
SQ	Sequence 32 AA:
Query Match	73.2%; Score 120; DB 15; Length 32;
Best Local Similarity	76.9%; Pred. No. 2, 5e-10;
Matches 20; Conservative 3; Mismatches 3; Indels 0; Gaps 0;	
Df	
	1 PEMSKECYQMRRMKLGAPSTTCRR 26
	:: :: :: :: :: :: ::
	5 PEMFKCRWMRMKKLGAPSTTCVRR 30
RESULT 15	
ID	AAR57466
	AAR57466 standard; Protein; 32 AA.
AC	AAR57466;
XX	
DT	28-FEB-1995 (first entry)
XX	
DE	Lactoferrin derived peptide #29.
XX	
KW	Lactoferrin, chemical; enzymatic; hydrolysis; antimicrobial;
KW	antiseptic; ischaemic disease.
OS	Mus musculus.
XX	
PN	JP06172200-A.
XX	
PD	21-JUN-1994.
XX	
PF	08-DEC-1992; 92JP-0327738.
PR	08-DEC-1992; 92JP-0327738.
XX	
PA	(MORG) MORINAGA MILK IND CO LTD.
XX	
WI	WPI; 1994-238662/29.
DR	
XX	
PT	Brain protectant for preventing ischaemic diseases without side
PT	effects - comprising 31 specified peptide(s), prepd. by
PT	lactoferrin hydrolysis
DS	Disclosure; Page 10; 11pp; Japanese.

XX The sequences given in AAR57438-68 represent fragments of lactoferrin
CC which were derived from the full length protein by chemical or enzyme
CC hydrolysis. These peptides have brain protecting properties, as
CC well as anti-microbial activity. Compositions containing these
CC peptides may be prepared with out the addition of antiseptics, and
CC may be administered at doses of at least 10 mg for parenteral
CC administration and 100 mg for oral administration. These peptides
CC are stable, heat resistant, water soluble and may be used for the
CC prevention of ischemic diseases without side effects.
XX

SQ Sequence 32 AA;

Query Match 73.2%; Score 120; DB 15; Length 32;
Best Local Similarity 76.9%; Pred. No. 2.5e-10;
Matches 20; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 PEMSKCYOMORRMKLGAPSTICRR 26
Db 5 PEMFKRRQWRMKLGAPSTICRR 30

Search completed: July 30, 2003, 16:23:28
Job time : 58.1369 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 49.1964 Seconds

(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-6

Perfect score: 164

Sequence: 1 PEMSKCYQWQRRNRKLGAPSTICIRRTSA 29

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_23:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*
16: sp_ivirus:*
17: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	61.0	234	4	Q8IX02 homo sapien
2	99	60.4	234	4	Q8IX02 homo sapien
3	97	59.1	711	4	Q8IX02 homo sapien
4	97	59.1	711	4	Q8IX02 homo sapien
5	97	59.1	711	4	Q8IX02 homo sapien
6	97	59.1	711	4	Q8IX02 homo sapien
7	97	59.1	711	4	Q8IX02 homo sapien
8	97	59.1	711	4	Q8IX02 homo sapien
9	97	59.1	711	4	Q8IX02 homo sapien
10	97	59.1	711	4	Q8IX02 homo sapien
11	97	59.1	711	4	Q8IX02 homo sapien
12	97	59.1	711	4	Q8IX02 homo sapien
13	97	59.1	711	4	Q8IX02 homo sapien
14	97	59.1	711	4	Q8IX02 homo sapien
15	97	59.1	711	4	Q8IX02 homo sapien
16	97	59.1	711	4	Q8IX02 homo sapien

17	53	32.3	1449	12	Q65974	Q65974 casava com
18	52	31.7	1399	11	Q8R011	Q8R011 mus musculu
19	51.5	31.4	1795	4	Q14674	Q14674 homo sapien
20	51.5	31.4	4437	2	Q8G3P3	Q8G3P3 streptomyc
21	51	31.1	489	16	P96223	P96223 mycobacteri
22	50	30.5	54	6	Q9TQV7	Q9TQV7 equus caball
23	50	30.5	154	16	Q8PF84	Q8PF84 xanthomonas
24	50	30.5	511	11	Q8BVH0	Q8BVH0 mus musculu
25	50	30.5	522	5	Q9WVL4	Q9WVL4 drosophila
26	50	30.5	643	11	Q8BV16	Q8BV16 mus musculu
27	50	30.5	1207	4	Q8TB59	Q8TB59 homo sapien
28	49.5	30.2	1536	4	Q9C014	Q9C014 homo sapien
29	49	29.9	130	17	Q8TVE4	Q8TVE4 methanopyru
30	49	29.9	195	12	Q91W22	Q91W22 hydra angu
31	49	29.9	258	5	Q9VAF6	Q9VAF6 drosophila
32	49	29.9	405	16	Q9ZAF8	Q9ZAF8 listeria in
33	49	29.9	408	16	Q8YAE4	Q8YAE4 listeria in
34	49	29.9	507	16	Q8A7G8	Q8A7G8 caulobacter
35	49	29.9	558	2	Q936U3	Q936U3 rhodospirill
36	49	29.9	1721	5	Q93547	Q93547 caenorhabdi
37	48.5	29.6	71	16	Q8XJ54	Q8XJ54 yersinia pe
38	48.5	29.6	132	16	Q8CLQ9	Q8CLQ9 yersinia pe
39	48.5	29.6	518	16	Q911H4	Q911H4 streptomyce
40	48	29.3	271	16	Q98JPH	Q98JPH rhizobium l
41	48	29.3	352	7	Q9XS13	Q9XS13 althocara
42	48	29.3	352	16	Q8XQ97	Q8XQ97 talstonia s
43	48	29.3	353	10	Q9ZTF0	Q9ZTF0 oryza sativ
44	48	29.3	402	10	Q9ZRH8	Q9ZRH8 oryza sativ
45	48	29.3	424	16	Q98FV3	Q98FV3 rhizobium l

ALIGNMENTS

RESULT 1
ID Q8IX02 PRELIMINARY; PRT; 234 AA.
AC Q8IX02;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Lactoferrin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu Z., Li N.,
RT "Human lactoferrin gene, 5' regulation region and exons 1-6";
RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF508798; AAO14686.1; -
FT NON_TER 234
SQ SEQUENCE 234 AA; 25671 MW; 2A08DAFD6B8780C7 CRC64;

Query Match 61.0%; Score 100; DB 4; Length 234;
Best Local Similarity 57.1%; Pred. No. 64-07; Mismatches 6; Gaps 0;
Matches 16; Conservative 6; Indels 0;

QY 1 PEMSKCYQWQRRNRKLGAPSTICIRRTS 28
DB 34 PEATKCFQWQRRNRKRGVPSICIRKDS 61
RESULT 2
ID Q8TCD2 PRELIMINARY; PRT; 711 AA.
AC Q8TCD2;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Lactoferrin.
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OK NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RA Strausberg R.;
 RT Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: TRANSFERRIN ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC DR EMBL; BC022347; AA422347.1; -
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR SMART; SMO0094; TR.PER. 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 2.
 KW Glycoprotein; Iron transport; Metal-binding; Transport.
 SQ SEQUENCE 711 AA; 78327 MW; 1B9C7EE097C45FAF CRC64;

Query Match 60.4%; Score 99; DB 4; Length 711;
 Best Local Similarity 57.1%; Pred. No. 2.5e-06;
 Matches 16; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMSKCYQWQNRKLGAPSTICIRTS 28
 DB 34 PEATKCFQWQNRKRVGPPVSCIKRDS 61

RESULT 3

ID 081ZHG PRELIMINARY; PRT; 711 AA.

AC 081ZHG;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Lactoferrin.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OK NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kaplan J.B.; Fine D.H.;
 RT "Characterization of an amino acid polymorphism in the antibacterial
 domain of human lactoferrin";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY137470; AA11394.1; -
 SQ SEQUENCE 711 AA; 78396 MW; 547AB9423C7C67 CRC64;

Query Match 59.1%; Score 97; DB 4; Length 711;
 Best Local Similarity 53.6%; Pred. No. 5.1e-06;
 Matches 15; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMSKCYQWQNRKLGAPSTICIRTS 28
 DB 34 PEATKCFQWQNRKRVGPPVSCIKRDS 61

RESULT 4

ID 08IU92 PRELIMINARY; PRT; 711 AA.

AC 08IU92;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Lactotransferrin precursor (Lactoferrin).
 GN LTP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OK NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Shi Y.-Q.; Zhang Y.; Zheng Y.-M.;
 RT "Homo sapiens Lactotransferrin Gene: cDNA Cloning and Sequence
 Analysis";
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 CC [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Seminal vesicle;
 RA Baekar Singh S.; Saravanan K.; Paramesivam M.; Srinivasan A.;
 RA Singh T.P.;
 RT "Homo sapiens Lactoferrin (HLF) mRNA";
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY165046; AA6398.1; -
 DR EMBL; AY178998; AA75578.2; -
 KW Signal.
 FT SIGNAL.
 FT CHAIN.
 SQ SEQUENCE 711 AA; 78382 MW; 547BF42C9267E67 CRC64;

Query Match 59.1%; Score 97; DB 4; Length 711;
 Best Local Similarity 53.6%; Pred. No. 5.1e-06;
 Matches 15; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMSKCYQWQNRKLGAPSTICIRTS 28
 DB 34 PEATKCFQWQNRKRVGPPVSCIKRDS 61

RESULT 5

ID 09TR80 PRELIMINARY; PRT; 33 AA.

AC 09TR80;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE Lactoferrin (Fragment).
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 OK NCBI_TaxID=9940;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=95127729; Pubmed=7827104;
 RA Qian Z.Y.; Jolles P.; Migliore-Samouir D.; Fiat A.M.;
 RL Biochim. Biophys. Acta 1243:25-32(1995).
 DR HSRP; O77698; ICE2
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 1.
 SQ SEQUENCE 33 AA; 3914 MW; D1904CAE15A73961 CRC64;

Query Match 55.5%; Score 91; DB 6; Length 33;
 Best Local Similarity 78.9%; Pred. No. 1.9e-06;
 Matches 15; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PMSKCYQWQNRKLGAP 19
 DB 14 PEGSKCYQWQNRKLGAP 32

RESULT 6

ID 08CBA0 PRELIMINARY; PRT; 707 AA.

AC 08CBA0;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Lactotransferrin.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Bone;
RX MEDLINE=22354683; PubMed=12466851;
RA THE FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
  60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK036451; BAC29450.1;
SQ SEQUENCE 707 AA; 77837 MW; E1B32F5FD8748A0F CRC64;

Query Match
Best Local Similarity 44.4%; Score 84; DB 11; Length 707;
Matches 12; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 EMSKCYQWRMRKLGAPSLTCIRRTS 28
Db 33 EEEKCLRMQNMKRVGGPPLSCVKSS 59

RESULT 7
OSMNB8 PRELIMINARY; PRT; 704 AA.
ID Q8MNB8 01-MAR-2002 (TREMBlrel. 20, Created)
AC Q8MNB8 01-MAR-2002 (TREMBlrel. 20, last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)
DE Lactoferrin (Fragment).
GN LTF.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RT Wang S.-R., Lin T.-Y., Wang C.-N.;
RT "Isolation and expression of porcine milk lactoferrin.";
RT Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TRANSFERRIN ARE IRON BINDING TRANSPORT PROTEINS WHICH
  CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
  OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC EMBL; L77887; AAL40161.1;
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SM00094; TR_FER; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 1.
KW Glycoprotein; Iron transport; Metal-binding; Transport.
FT NON TER
SQ SEQUENCE 704 AA; 77681 MW; 64EB769F7503CC32 CRC64;

Query Match
Best Local Similarity 51.9%; Score 69; DB 6; Length 704;
Matches 14; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

QY 2 EMSKCYQWRMRKLGAPSLTCIRRTS 28
Db 34 EYSKCRQWQSKIRRTN--PIFCIRRA 58

RESULT 8
OSMNB8 PRELIMINARY; PRT; 38 AA.
ID Q9UCY5 01-MAY-2000 (TREMBlrel. 13, Created)
AC Q9UCY5 01-MAY-2000 (TREMBlrel. 13, last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)

```

```

DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)
DE Lactoferrin homolog (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=96081613; PubMed=8551695;
RT Sato I.;
RT "Characterization of the 84-kDa protein with ABH activity in human
  seminal plasma.";
RT Jpn. J. Legal Med. 43:281-293(1995).
DR HSSP; 077811; 1B1X.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 1.
SQ SEQUENCE 38 AA; 4459 MW; 0402F49B5EBDDB CRC64;

Query Match
Best Local Similarity 41.5%; Score 68; DB 4; Length 38;
Matches 12; Conservative 50.0%; Pred. No. 0.006; Mismatches 8; Indels 0; Gaps 0;

QY 1 PEWSKCYQWRMRKLGAPSLTCI 24
Db 15 PEADKXFWQNMKRVGGPPLSCVKSS 38

RESULT 9
ID Q9DBD0 PRELIMINARY; PRT; 700 AA.
AC Q9DBD0 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DE 1300017J02Rik protein.
DE 1300017J02Rik.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Toch M., Ishii Y.,
RA Aizawa K., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Giesi C., King B., Kochiya H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Watanabe T.,
RA Sakai K., Okido T., Futuro M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Butt C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima T., Mazzarelli J., Mombert P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seya T., Shihata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Welz C., Whitteker C., Wilming L.,
RA Wymshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
DR EMBL; AK005035; BAB23762.1;
DR HSSP; P19134; 1TFD.
DR MGD; MGI:1919025; 1300017J02Rik.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SM00094; TR_FER; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 1.
DR PROSITE; PS00207; TRANSFERRIN_3; 1.

```

SQ SEQUENCE 700 AA; 76765 MW; 38C991D1021AB548 CRC64;

Query Match 36.9%; Score 60.5; DB 11; Length 700;
 Best Local Similarity 36.7%; Pred. No. 1.4;
 Matches 11; Conservative 9; Mismatches 7; Indels 3; Gaps 1;

OY 2 EWSKCYOWRMRKLT---GAPSTICIRRTS 28
 |||||::|||:|||||:|||||:|
 DB 34 EATKCSFRDNMKVLPAGGPATVCRKMS 63

RESULT 10

O9XT72 PRELIMINARY; PRT; 711 AA.

AC O9XT72;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Serum iron transport protein transferrin.
 GN TRFE.
 OS Trichosurus vulpecula (Brush-tailed possum).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
 OX NCBI_TaxID=9337;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=99227120; PubMed=10209259;
 RA Demmer J., Stasik S.J., Adamski F.M., Grigor M.R.;
 RT "Cloning and expression of the transferrin and ferritin genes in a
 RT marsupial, the brush-tailed possum (Trichosurus vulpecula).";
 RL Biochim. Biophys. Acta 1445:65-74(1999).
 CC -1- FUNCTION: TRANSFERRING ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION. USUALLY BICARBONATE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC DR EMBL; AF092510; AAD58311.1; -.
 DR HSP; P56410; IAOV.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PRO0422; TRANSFERRIN.
 DR SMART; SMO0094; TR_FER; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 1.
 DR PROSITE; PS00207; TRANSFERRIN_3; 1.
 DR Glycoprotein; iron transport; Metal-binding; Transport.
 SQ SEQUENCE 711 AA; 77764 MW; 35CA494D95631F CRC64;

Query Match 35.4%; Score 58; DB 6; Length 711;
 Best Local Similarity 33.3%; Pred. No. 3.4;
 Matches 9; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

OY 2 EWSKCYOWRMRKLTGAPSTICIRRTS 28
 |||||::|||:|||||:|||||:|
 DB 33 ELKCKNEFKEMNSINHPGLACIKTIS 59

RESULT 11

O9VC96 PRELIMINARY; PRT; 700 AA.

AC O9VC96;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE RIKEN cDNA 1300017J02 gene.
 GN 1300017J02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Strusberg R.;

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC021390; AA021390.1; -.
 DR MGI; MGI:1918025; 1300017J02RIK.
 DR InterPro; IPR001156; transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PRO0422; TRANSFERRIN.
 DR SMART; SMO0094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 1.
 DR PROSITE; PS00207; TRANSFERRIN_3; 1.
 DR SEQUENCE 700 AA; 76795 MW; 7DA4A560CBB6BF37 CRC64;

Query Match 35.1%; Score 57.5; DB 11; Length 700;
 Best Local Similarity 37.0%; Pred. No. 4;
 Matches 10; Conservative 8; Mismatches 6; Indels 3; Gaps 1;

OY 5 KCYOWRMRKLT---GAPSTICIRRTS 28
 |||||::|||:|||||:|||||:|
 DB 37 KCSSFRDNMKVLPAGGPATVCRKMS 63

RESULT 12

O8KL39 PRELIMINARY; PRT; 192 AA.

AC O8KL39;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Hypothetical protein.
 GN YP044.
 OS Rhizobium etli.
 OG Plasmid symbiotic plasmid p42d.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=28449;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CFN42;
 RX MEDLINE=91193195; PubMed=2013564;
 RA Girard M.L., Flores M., Brom S., Romero D., Palacios R., Davila G.;
 RT "Structural complexity of the symbiotic plasmid of Rhizobium
 RT leguminosarum bv. phaseoli.";
 RL J. Bacteriol. 173:2411-2419(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CFN42;
 RX MEDLINE=97419521; PubMed=9274036;
 RA Ramirez-Romero M.A., Buscos P., Girard L., Rodriguez O.,
 RA Cevallos M.A., Davila G.;
 RT "Sequence, localization and characteristics of the replicator region
 RT of the symbiotic plasmid of Rhizobium etli.";
 RL Microbiology 143:2825-2831(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CFN42;
 RA Quintero V., Cevallos M.A., Davila G.;
 RT "A site-specific recombinase and RecA are required to exert
 RT incompatibility towards the symbiotic plasmid of Rhizobium etli.";
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U80928; AAM54928.1; -.
 KM Hypothetical protein; Plasmid.
 SQ SEQUENCE 192 AA; 21513 MW; 7AC7624DFOEB232E CRC64;

Query Match 33.2%; Score 54.5; DB 2; Length 192;
 Best Local Similarity 60.0%; Pred. No. 3.1;
 Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

OY 9 WQRMRLGAPSTICIRRTS 28
 |||||::|||:|||||:|||||:|
 DB 148 WQERQSGAP-VTAIRNS 166

RESULT 13

O8KPR3

```

ID 08KPR9 PRELIMINARY; PRT; 821 AA.
AC 08KPR9;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein.
GN SEB0025.
OS Synecchococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAINPCC 7492;
RA Holtman C.K., Sandoval P., Chen Y., Socias T., Mohler B.J.,
RA McMurtry S., Gonzalez A., Salinas I., Golden S.S., Youderlian P.;
RA "Synecchococcus elongatus PCC7942 cosmid 7G3."
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY120853; AAM82700.1; -.
SQ Hypothetical protein
KW SEQUENCE 821 AA; 91706 MW; D5221A0D0ED0DBCC CRC64;
SQ

Query Match 32.3%; Score 54; DB 2; Length 821;
Best Local Similarity 47.8%; Pred. No. 16;
Matches 11; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 PEMSKCYQWORMRKLGPSTIC 23
Db 517 PDSQALWQBRALTTSAKPTIC 539

RESULT 14
097490 PRELIMINARY; PRT; 736 AA.
AC 097490;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE P97.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98451505; Pubmed=9780225;
RA Kawamoto T., Pan H., Yan W., Ishida H., Usui E., Oda R., Nakamasu K.,
RA Noshito M., Kawashima-Ohya Y., Fujii M., Shintani H., Okada Y.,
RA Kato Y.;
RT "Expression of membrane-bound transferrin-like protein p97 on the cell
RT surface of chondrocyte."
RU Eur. J. Biochem. 256:503-509(1998).
CC -I- FUNCTION: TRANSFERRIN ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
DR EMBL; AB010995; BAA33956.1; -.
DR HSSP; P19134; ITED.
DR InterPro; IPR001064; CrystalIn.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; Transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SM00094; TR_FER; 2.
DR PROSITE; PS00225; CRYSTALININ_BETAGAMMA; 1.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 1.
DR PROSITE; PS00207; TRANSFERRIN_3; 1.
DR GlycoProtein; Iron transport; Metal-binding; Transport.
KW SEQUENCE 736 AA; 80169 MW; F389D9F8A6AC90FC CRC64;
SQ

Query Match 32.6%; Score 53.5; DB 6; Length 736;
Best Local Similarity 36.7%; Pred. No. 17;
Matches 11; Conservative 6; Mismatches 12; Indels 1; Gaps 1;

```

```

QY 1 PEMSKCYQWORMRKLGPSTICIRPTSA 29
Db 31 PEQKCEDMSQAFRAGIQPALICVQGTSA 60

RESULT 15
077558 PRELIMINARY; PRT; 48 AA.
AC 077558;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Lactoferrin (fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98332734; Pubmed=9666128;
RA Wang S.R., Lin J., Cheng I.C., Lin T.Y.;
RT "Characterization and functional analysis of the porcine lactoferrin
RT gene promoter."
RL Gene 215:203-212 (1998).
DR EMBL; AF044256; AAC34369.1; -.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; Transferrin; 1.
FT NON_TER 48
SQ SEQUENCE 48 AA; 5569 MW; 28403BETDE144D78 CRC64;
SQ

Query Match 32.3%; Score 53; DB 6; Length 48;
Best Local Similarity 57.1%; Pred. No. 1.3;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 EMSKCYQWORMRKL 15
Db 35 EYSKCRQWQSKRR 48

```

Search completed: July 30, 2003, 16:29:25
Job time : 49.1964 secs

THIS PAGE BLANK (user10)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 30, 2003, 16:10:10 ; Search time 9.49405 Seconds
(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-6

Sequence: 1 PEMSKCYQMQRMRKLGAPSTICIRRTSA 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	164	100.0	708	TRPL_CAPHI	Q29477 capra hircu
2	124	75.6	708	TRPL_BUBBU	O77698 bubalus bub
3	122	74.4	708	TRPL_BOVIN	P24627 bos taurus
4	104	63.4	708	TRPL_CAMDR	Q94um0 camelus dro
5	100	61.0	711	TRPL_HUMAN	P02788 homo sapien
6	84	51.2	707	TRPL_MOUSE	P08071 mus musculu
7	83	50.6	695	TRPL_HORSE	O77811 equus caball
8	66	40.2	704	TRPL_PIG	P14632 sus scrofa
9	58.5	35.7	704	ICA_FIG	Q29545 sus scrofa
10	58	35.4	271	MEGP_ECO57	Q8xb95 escherichia
11	58	35.4	271	MEGP_ECOLI	P76329 escherichia
12	58	35.4	271	MEGP_SALTY	Q8xge1 salmonella
13	57.5	35.1	738	TRFM_HUMAN	P06582 homo sapien
14	52	31.7	198	PEMT_MOUSE	Q61907 mus musculu
15	52	31.7	271	MEGP_ECOLI6	P53286 escherichia
16	51.5	31.4	1795	ESPI_HUMAN	Q14674 homo sapien
17	51	31.1	198	PEMT_RAT	Q08388 rattus norv
18	50	30.5	642	2398_HUMAN	Q8rd17 homo sapien
19	50	30.5	706	TRFE_HORSE	P27425 equus caball
20	50	30.5	1207	AT19_HUMAN	Q8te59 homo sapien
21	48	29.3	430	SYH_CHLNP	Q947p1 chlamydia p
22	47.5	29.0	450	PEM2_RAT	Q94vvt rattus norv
23	47	28.7	198	PEMT_HUMAN	Q9ubm1 homo sapien
24	47	28.7	704	TRAL_HUMAN	Q12931 homo sapien
25	47	28.7	2671	IP3T_HUMAN	Q14573 homo sapien
26	46.5	28.4	695	TRFE_RABIT	P13134 onyctolagus
27	46	28.0	63	BD03_MOUSE	Q9wcl0 mus musculu
28	46	28.0	413	YEIT_SALTI	Q8z546 salmonella
29	46	28.0	413	YEIT_SALTY	Q8zn18 salmonella
30	46	28.0	488	VE2_HPV49	Q96795 human papil
31	46	28.0	647	TOP3_VIBCH	Q9q6fs vibrio chol
32	46	28.0	1082	A3B2_HUMAN	Q13367 homo sapien
33	46	28.0	1210	AT19_MOUSE	P53509 mus musculu

34	45.5	27.7	641	1	SCAB_RABIT	O97747 onyctolagus
35	45.5	27.7	698	1	TRFE_HUMAN	P02788 homo sapien
36	45.5	27.7	967	1	ATSI_HUMAN	Q9uh18 homo sapien
37	45	27.4	199	1	NHAA_PEBCL	P27764 pseudomonas
38	45	27.4	310	1	TF2B_METTH	Q26974 methanobact
39	45	27.4	368	1	HA14_MOUSE	P14427 mus musculu
40	45	27.4	1385	1	RPPD_PLAWV	Q07518 plantago as
41	45	27.4	1490	1	CRK7_HUMAN	Q9nyv4 homo sapien
42	44.5	27.1	937	1	SYL_METTH	Q27552 methanobact
43	44.5	27.1	1793	1	YCPI_LOTJA	Q9bne6 lotus japon
44	44	26.8	142	1	RPPD_OENEB	P18493 oenothera d
45	44	26.8	236	1	YHIL_ECOLI	P31213 escherichia

ALIGNMENTS

RESULT 1
TRPL_CAPHI STANDARD; PRT; 708 AA.
AC Q29477; Q29479;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactotransferrin precursor (lactoferrin).
GN LTF.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Lee T., Yu S., Kim S., Lee K., Yu D.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RT [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=94380047; PubMed=8093048;
RT "le Provost F., Nocart M., Guerin G., Martin P.;
RT "Characterization of the goat lactoferrin cDNA. Assignment of the
RT "relevant locus to bovine U12 synteny group.";
RT Biochem. Biophys. Res. Commun. 203:1324-1332(1994).
CC -!- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC OF AN ANION. USUALLY BICARBONATE.
CC -!- SUBUNIT: Monomer (by similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U53857; AAA97958.1; -
CC EMBL: X78902; CAAS5517.1; -
CC HSSP: O77698; ICE2.
DR InterPro: IPR001156; Transferrin.
DR Pfam: PF00405; transferrin; 2.
DR PRINTS: PR00422; TRANSFERRIN.
DR SMART: SM00394; TR_FER; 2.
DR PROSITE: PS00205; TRANSFERRIN_1; 2.
DR PROSITE: PS00206; TRANSFERRIN_2; 2.
DR PROSITE: PS00207; TRANSFERRIN_3; 2.
KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
KW Signal.
FT SIGNAL 1 19 BY SIMILARITY.

FT CHAIN 20 708 LACTOTRANSFERRIN.
 FT REPEAT 20 363 1.
 FT DISULFID 364 708 2. BY SIMILARITY.
 FT DISULFID 28 64 BY SIMILARITY.
 FT DISULFID 38 55 BY SIMILARITY.
 FT DISULFID 134 217 BY SIMILARITY.
 FT DISULFID 176 192 BY SIMILARITY.
 FT DISULFID 189 200 BY SIMILARITY.
 FT DISULFID 250 264 BY SIMILARITY.
 FT DISULFID 367 399 BY SIMILARITY.
 FT DISULFID 377 390 BY SIMILARITY.
 FT DISULFID 424 703 BY SIMILARITY.
 FT DISULFID 444 666 BY SIMILARITY.
 FT DISULFID 476 551 BY SIMILARITY.
 FT DISULFID 500 694 BY SIMILARITY.
 FT DISULFID 510 524 BY SIMILARITY.
 FT DISULFID 521 534 BY SIMILARITY.
 FT DISULFID 592 606 BY SIMILARITY.
 FT DISULFID 644 649 BY SIMILARITY.
 FT METAL 79 79 BY SIMILARITY.
 FT METAL 111 111 BY SIMILARITY.
 FT METAL 211 211 BY SIMILARITY.
 FT METAL 272 272 BY SIMILARITY.
 FT METAL 414 414 BY SIMILARITY.
 FT METAL 452 452 BY SIMILARITY.
 FT METAL 545 545 BY SIMILARITY.
 FT METAL 614 614 BY SIMILARITY.
 FT BINDING 140 140 ANION (BY SIMILARITY).
 FT BINDING 482 482 ANION (BY SIMILARITY).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 56 56 I -> V (IN REF. 2).
 FT CONFLICT 88 88 Q -> K (IN REF. 2).
 FT CONFLICT 124 124 Q -> P (IN REF. 2).
 FT CONFLICT 154 154 S -> R (IN REF. 2).
 FT CONFLICT 304 304 S -> G (IN REF. 2).
 FT CONFLICT 414 414 D -> G (IN REF. 2).
 FT SEQUENCE 708 AA; 77358 MM; P2EDA3C8353960D CRC64;

Query Match Best Local Similarity 100.0%; Score 164; DB 1; Length 708;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PEWSKCYQWORMRKLAGPSITCIRRTSA 29
 DB 33 PEWSKCYQWORMRKLAGPSITCIRRTSA 61

RESULT 2
 TRFL_BUBBU STANDARD; PRT; 708 AA.
 ID TRFL_BUBBU
 AC 07698;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Lactotransferrin precursor (lactoferrin).
 GN LTP
 OS Bubalus bubalis (Domestic water buffalo).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bubalus.
 NCBI_TaxID=89462;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Paramesivam M., Thattaiyath B.D., Kumar A., Srinivasan A.,
 RA Singh T.P.;
 RL "cDNA sequence of Buffalo lactoferrin."
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).

RX MEDLINE=20003130; PubMed=10531476;
 RA Kathirkeyan S., Paramesivam M., Yadav S., Srinivasan A., Singh T.P.;
 RT Structure of buffalo lactoferrin at 2.5-A resolution using crystals
 grown at 303 K shows different orientations of the N and C lobes.
 RL Acta Crystallogr. D 55:1805-1813(1999).
 CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 OF AN ANION, USUALLY BICARBONATE.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL, AJ005203; CA006441.1; .
 CC PDB; 1CE2; 19-MAR-99.
 CC PDB; 1B1Y; 13-JAN-99.
 CC InterPro; IPRO01156; Transferrin.
 CC Pfam; PF00405; transferrin; 2.
 CC PRINTS; PR00422; TRANSFERRIN.
 CC SMART; SM00094; TR_FBR; 2.
 CC PROSITE; PS00205; TRANSFERRIN_1; 2.
 CC PROSITE; PS00206; TRANSFERRIN_2; 2.
 CC PROSITE; PS00207; TRANSFERRIN_3; 2.
 CC Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 CC Signal; 3d-structure.
 CC CHAIN 1 19
 FT DISULFID 28 708 LACTOTRANSFERRIN.
 FT DISULFID 38 64
 FT DISULFID 134 217
 FT DISULFID 176 192
 FT DISULFID 179 202
 FT DISULFID 189 200
 FT DISULFID 250 264
 FT DISULFID 367 399
 FT DISULFID 377 390
 FT DISULFID 424 703
 FT DISULFID 444 666
 FT DISULFID 476 551
 FT DISULFID 500 694
 FT DISULFID 510 524
 FT DISULFID 521 534
 FT DISULFID 592 606
 FT DISULFID 644 649
 FT METAL 79 79 IRON 1.
 FT METAL 111 111 IRON 1.
 FT METAL 211 211 IRON 1.
 FT METAL 272 272 IRON 1.
 FT METAL 414 414 IRON 2.
 FT METAL 452 452 IRON 2.
 FT METAL 545 545 IRON 2.
 FT METAL 614 614 IRON 2.
 FT BINDING 140 140 ANION (POTENTIAL).
 FT BINDING 482 482 ANION (POTENTIAL).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 21 22
 FT STRAND 25 25
 FT HELIX 32 46
 FT TURN 47 48
 FT STRAND 53 57
 FT HELIX 61 69
 FT TURN 70 71

```

FT STRAND 75 78
FT HELIX 80 87
FT TURN 89 91
FT STRAND 93 102
FT STRAND 107 108
FT STRAND 110 118
FT TURN 119 120
FT HELIX 125 127
FT TURN 129 130
FT STRAND 132 135
FT TURN 138 139
FT TURN 141 144
FT HELIX 145 150
FT TURN 151 151
FT HELIX 152 155
FT TURN 159 161
FT HELIX 164 169
FT TURN 170 171
FT STRAND 174 176
FT TURN 178 179
FT TURN 182 184
FT HELIX 186 189
FT TURN 190 191
FT HELIX 196 198
FT TURN 199 200
FT TURN 204 205
FT TURN 207 208
FT HELIX 211 219
FT TURN 220 221
FT STRAND 225 229
FT TURN 230 231
FT HELIX 232 236
FT HELIX 240 243
FT TURN 244 245
FT STRAND 246 249
FT TURN 251 252
FT STRAND 255 257
FT HELIX 258 260
FT TURN 261 263
FT STRAND 267 270
FT STRAND 273 277
FT HELIX 283 297
FT TURN 299 300
FT TURN 307 308
FT TURN 312 313
FT TURN 321 322
FT STRAND 325 328
FT TURN 331 332
FT HELIX 335 351
FT STRAND 354 362
FT STRAND 364 369
FT HELIX 371 384
FT TURN 385 385
FT STRAND 388 393
FT HELIX 396 404
FT TURN 405 406
FT STRAND 410 413
FT HELIX 415 422
FT TURN 423 425
FT STRAND 426 434
FT TURN 438 441
FT TURN 444 446
FT STRAND 452 459
FT TURN 460 461
FT TURN 463 464
FT HELIX 466 469
FT TURN 471 472
FT STRAND 474 477
FT TURN 480 481
FT TURN 483 486
FT HELIX 487 497
FT TURN 503 505
FT STRAND 508 510

```

```

FT TURN 512 513
FT TURN 516 517
FT TURN 519 520
FT TURN 522 523
FT TURN 528 529
FT TURN 533 534
FT TURN 538 539
FT TURN 541 542
FT HELIX 544 553
FT TURN 554 555
FT STRAND 559 563
FT HELIX 564 569
FT TURN 570 570
FT HELIX 578 581
FT TURN 582 582
FT TURN 585 586
FT STRAND 588 591
FT TURN 593 594
FT STRAND 597 599
FT TURN 600 601
FT TURN 603 605
FT STRAND 609 611
FT STRAND 615 619
FT HELIX 620 637
FT TURN 639 640
FT TURN 642 647
FT TURN 650 651
FT TURN 654 655
FT TURN 662 663
FT STRAND 664 669
FT HELIX 676 680
FT HELIX 682 692
FT TURN 693 694
FT HELIX 698 705

```

```

Query Match 75.6%; Score 124; DB 1; Length 708;
Best Local Similarity 76.9%; Pred. No. 1.2e-10;
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 PEMSKCYOMORMRKLGAPSTICRR 26
Db 33 PEMKCHRMQMRKKGAPSTICVR 58

```

```

RESULT 3
TRFL_BOVIN
ID TRFL_BOVIN STANDARD; PRT; 708 AA.
AC P24627; Q29629; Q9WZV3;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactoferrin precursor (Lactoferrin) [Contains: Lactoferricin B
  (LFcIN B)].
GN LTF.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
[1]
SEQUENCE FROM N.A.
RC TISUB=Submaxillary gland;
RX MEDLINE=9116050; PubMed=2001696;
RA Pierce A., Colavizza D., Benaisa M., Maes P., Tartar A.,
  Montreuil J., Spik G.;
RT "Molecular cloning and sequence analysis of bovine lactoferrin.",
  Eur. J. Biochem. 196;177-184(1991).
[2]
SEQUENCE FROM N.A.
RP MEDLINE=92028986; PubMed=1718281;
RA Goodman R.E., Schanbacher F.L.;
RT "Bovine lactoferrin mRNA: sequence, analysis, and expression in the
  mammary gland.";

```

RL Biochem. Biophys. Res. Commun. 180:75-84(1991).
 RN [31] SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Tsang T.C., Burns D.K., Wang F., Pan Y.C.E., Schmidt A.M., Stern D.;
 RT "Cloning of a 80-kD advanced glycosylation end product (AGE) binding
 protein from bovine lung";
 RN FASEB J. 6:233-233(1991).
 [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood, and Mammary gland;
 RX MEDLINE=94266164; PubMed=8206385;
 RA Seyler H.-M., Tuckoritz A., Interthal H., Koczan D., Hobom G.;
 RT "Structure of the bovine lactoferrin-encoding gene and its promoter";
 RN Gene 143:265-269(1994).
 [5]
 RP SEQUENCE FROM N.A.
 RN Nakamura I., Shimazaki K., Yagi Y., Watanabe A.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 [6]
 RP SEQUENCE OF 20-59.
 RX MEDLINE=90031466; PubMed=2805645;
 RA Rejman J.J., Hegarty H.M., Hurley W.L.;
 RT "Purification and characterization of bovine lactoferrin from
 secretions of the involuting mammary gland: identification of
 multiple molecular weight forms";
 RN Comp. Biochem. Physiol. 93B:929-934(1989).
 [7]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE=98062367; PubMed=9398529;
 RA Moore S.A., Anderson B.F., Groom C.R., Haridas M., Baker E.N.;
 RT "Three-dimensional structure of dimeric bovine lactoferrin at 2.8-A
 resolution";
 RN J. Mol. Biol. 274:222-236(1997).
 [8]
 RP STRUCTURE BY NMR OF 36-60.
 RX MEDLINE=98190007; PubMed=9521752;
 RA Hwang P.M., Zhou N., Shan X., Arrowsmith C.H., Vogel H.J.;
 RT "Three-dimensional solution structure of lactoferricin B, an
 antimicrobial peptide derived from bovine lactoferrin";
 RN Biochemistry 37:4288-4298(1998).
 CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 OF AN ANION, USUALLY BICARBONATE.
 CC -1- FUNCTION: LACTOFERRICIN B IS AN ANTIMICROBIAL PEPTIDE.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, X57084, CAA40266.1, -;
 DR EMBL, M63502, AAA30617.1, -;
 DR EMBL, L08604, AAA30609.1, -;
 DR EMBL, L19993, AAA21722.1, -;
 DR EMBL, L19982, AAA21722.1, JOINED.
 DR EMBL, L19983, AAA21722.1, JOINED.
 DR EMBL, L19984, AAA21722.1, JOINED.
 DR EMBL, L19985, AAA21722.1, JOINED.
 DR EMBL, L19986, AAA21722.1, JOINED.
 DR EMBL, L19988, AAA21722.1, JOINED.
 DR EMBL, L19989, AAA21722.1, JOINED.
 DR EMBL, L19990, AAA21722.1, JOINED.
 DR EMBL, L19991, AAA21722.1, JOINED.
 DR EMBL, L19992, AAA21722.1, JOINED.
 DR EMBL, AB046664, BAB03470.1, -;

DR PIR, I45919, TPBOL.
 DR PDB, 1BLF, 03-DEC-97.
 DR PDB, 1LFC, 18-NOV-98.
 DR GlycoSuiteDB; P24627; -;
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00205; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 2.
 DR Transport; Iron transport; Glycoprotein, Metal-binding; Repeat;
 KW Signal; Antibiotic; 3D-structure.
 FT SIGNAL 1 19
 FT CHAIN 20 708
 FT PEPTIDE 36 60
 FT REPEAT 20 363
 FT REPEAT 364 708
 FT DISULFID 28 64
 FT DISULFID 38 55
 FT DISULFID 134 217
 FT DISULFID 176 192
 FT DISULFID 189 200
 FT DISULFID 250 264
 FT DISULFID 367 399
 FT DISULFID 377 390
 FT DISULFID 424 703
 FT DISULFID 444 666
 FT DISULFID 476 551
 FT DISULFID 500 694
 FT DISULFID 510 524
 FT DISULFID 521 534
 FT DISULFID 592 606
 FT DISULFID 644 649
 FT METAL 79 79
 FT METAL 111 111
 FT METAL 211 211
 FT METAL 272 272
 FT METAL 414 414
 FT METAL 452 452
 FT METAL 545 545
 FT METAL 614 614
 FT BINDING 140 140
 FT BINDING 482 482
 FT CARBOHYD 252 252
 FT CARBOHYD 387 387
 FT CARBOHYD 495 495
 FT CARBOHYD 564 564
 FT CONFLICT 63 63
 FT CONFLICT 66 67
 FT CONFLICT 145 145
 FT CONFLICT 164 165
 FT CONFLICT 264 264
 FT CONFLICT 273 273
 FT CONFLICT 281 281
 FT CONFLICT 291 291
 FT CONFLICT 297 297
 FT CONFLICT 340 340
 FT CONFLICT 418 418
 FT CONFLICT 439 439
 FT CONFLICT 459 459
 FT CONFLICT 514 514
 FT CONFLICT 632 632
 FT STRAND 25 29
 FT HELIX 32 45
 FT HELIX 46 48
 FT STRAND 53 57
 FT HELIX 61 69
 FT TURN 70 71
 FT STRAND 75 78
 E -> A (IN REF. 4).
 RA -> PG (IN REF. 2).
 I -> V (IN REF. 1 AND 4).
 LQ -> PP (IN REF. 1).
 C -> Y (IN REF. 4).
 A -> P (IN REF. 4).
 G -> A (IN REF. 4).
 S -> R (IN REF. 4).
 F -> S (IN REF. 2).
 I -> A (IN REF. 1).
 G -> V (IN REF. 1).
 H -> Y (IN REF. 1).
 K -> R (IN REF. 4).
 A -> R (IN REF. 1).
 H -> R (IN REF. 5).

```

FT HELIX 80 87
FT TURN 89 91
FT STRAND 93 102
FT STRAND 107 108
FT STRAND 110 118
FT TURN 126 127
FT TURN 129 130
FT STRAND 133 135
FT TURN 138 139
FT TURN 141 144
FT HELIX 145 155
FT TURN 156 156
FT TURN 159 161
FT HELIX 164 169
FT TURN 170 171
FT STRAND 175 176
FT TURN 178 179
FT TURN 182 184
FT HELIX 186 188
FT TURN 190 191
FT TURN 196 197
FT TURN 199 200
FT TURN 204 205
FT TURN 207 208
FT HELIX 210 219
FT TURN 220 221
FT STRAND 225 229

```

Query Match 74.4%; Score 122; DB 1; Length 708;
 Best Local Similarity 72.4%; Pred. No. 2.3e-10;
 Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

Qy 1 PEMSKCYQMRMRKLGAPSTICIRRTSA 29
Db 33 PEMFKCRMRMRKLGAPSTICVRRAFA 61

```

RESULT 4

TRFL CAMDR STANDARD; PRT; 708 AA.

```

AC Q9TUM0; Q9TUM0; 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 42, Last annotation update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lactoferrin precursor (Lactoferrin).
GN LTF.
OS Camelus dromedarius (Dromedary) (Arabian camel).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9938;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Small; TISSUE=lactating mammary gland;
RA Kasper S.R., Ackermann M., Farah Z., Puhon Z.;
RT "Sequence analysis of camel (Camelus dromedarius) lactoferrin";
RL Int. Dairy J. 9:481-486 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Paramasivam M., Srinivasan A., Singh R., Sahani M.S., Singh T.P.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.

```

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb.ch/announce/>
 or send an email to license@isb.ch).

```

CC EMBL; AJ131674; CAB53387.1; -.
CC EMBL; AF165879; AAF82241.1; -.
CC PDB; 1DTZ; 20-JUN-01.
CC InterPro; IPR001156; Transferrin.
CC Pfam; PF00405; transferrin; 2.
CC PRINTS; PR00422; TRANSFERRIN.
CC SMART; SM00094; TR_FER; 2.
CC PROSITE; PS00205; TRANSFERRIN_1; 2.
CC PROSITE; PS00206; TRANSFERRIN_2; 2.
CC PROSITE; PS00207; TRANSFERRIN_3; 2.
CC KEGG; Signal; 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 708
FT REPEAT 20 363
FT REPEAT 364 708
FT DISULFID 28 64
FT DISULFID 38 55
FT DISULFID 134 217
FT DISULFID 176 192
FT DISULFID 189 200
FT DISULFID 250 264
FT DISULFID 367 399
FT DISULFID 377 390
FT DISULFID 424 703
FT DISULFID 444 666
FT DISULFID 476 551
FT DISULFID 500 694
FT DISULFID 510 524
FT DISULFID 521 534
FT DISULFID 592 606
FT DISULFID 644 649
FT METAL 79 79
FT METAL 111 111
FT METAL 211 211
FT METAL 272 272
FT METAL 414 414
FT METAL 452 452
FT METAL 545 545
FT METAL 614 614
FT BINDING 140 140
FT BINDING 482 482
FT CARBOHYD 252 252
FT CARBOHYD 385 385
FT CARBOHYD 537 537
FT CARBOHYD 594 594
FT CONFLICT 261 261
FT CONFLICT 304 304
FT CONFLICT 330 330
FT CONFLICT 492 494
FT CONFLICT 506 506
FT CONFLICT 609 609
FT CONFLICT 642 642
SQ SEQUENCE 708 AA; 77211 MW; 0B0C175A0B69430 CAC64;

```

Query Match 63.4%; Score 104; DB 1; Length 708;
 Best Local Similarity 63.0%; Pred. No. 1.1e-07;
 Matches 17; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

```

Qy 2 EWSKCYQMRMRKLGAPSTICIRRTS 28
Db 34 EWSKCYQMRMRKLGAPSTICVTKTS 60

```

RESULT 5

TRFL HUMAN STANDARD; PRT; 711 AA.
 AC P02788; Q00756; Q16780; Q16785; Q16786; Q16789; Q96K24; Q96K25;
 AC Q9H123;

DT 21-JUL-1986 (Rel. 01, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Lactoferrin precursor (Lactoferrin) [Contains: Lactoferrin A,
Lactoferrin B; Lactoferrin C].

GN LTF OR LTF.

OS Homo sapiens (Human).

OC Bacteria; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Mammary gland;

RC MEDLINE=90384839; PubMed=2402455;

RA Rey M.W., Woloszyk S.L., de Boer H.A., Pieper F.R.;

RT "Complete nucleotide sequence of human mammary gland lactoferrin.";

RL Nucleic Acids Res. 18:5288-5288(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Mammary gland;

RC Cho Y.Y.;

RT Thesis (1994), Genetic Engineering Research Institute / Taejeon, Korea.

RN [3]

RP SEQUENCE FROM N.A.

RA Connely O.M.;

RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RC TISSUE=Mammary gland;

RC Liang Q., Jimenez-Flores R., Richardson T.;

RT "Molecular cloning and sequence analysis of human lactoferrin.";

RL Submitted (DEC-1991) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RC TISSUE=Bone marrow;

RC Wei X., Han J., Rado T.A.;

RT "Human neutrophil lactoferrin coding and 5' flanking region DNA
sequences.";

RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.

RN [6]

RP SEQUENCE FROM N.A.

RC TISSUE=Mammary gland;

RC Cheng H., Chen X., Huan L.;

RT "CDN cloning and sequence analysis of human lactoferrin.";

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

RN [7]

RP SEQUENCE FROM N.A.

RC TISSUE=Prostate;

RC MEDLINE=22388257; PubMed=12477932;

RA Strussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.L., Wang J., Hsieh F.,
Diachenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullany S.J.,
Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunnarsson P.H.,
Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahy J., Helton E., Kettman M., Madan A.C., Rodighiero S., Sanchez A.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield A.S.N., Kravinsky M.I., Skalska U., Smailus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [8]

RP SEQUENCE OF 3-711 FROM N.A.

RC TISSUE=Mammary gland;

RC MEDLINE=90326549; PubMed=2374734;

RA Powell M.J., Ogden J.E.;

RT "Nucleotide sequence of human lactoferrin cDNA.";

RL Nucleic Acids Res. 18:4013-4013(1990).

RN [9]

RP SEQUENCE OF 20-711.

RA MEDLINE=8507667; PubMed=6510420;

RA Metz-Boutigue M.-H., Jolles J., Mazurier J., Schoentgen F.,
Legrand D., Spik G., Montreuil J., Jolles P.;

RT "Human lactoferrin: amino acid sequence and structural
comparisons with other transferrins.";

RL Eur. J. Biochem. 145:659-666(1984).

RN [10]

RP PRELIMINARY SEQUENCE OF 20-73, 134-171, 257-278 AND 347-530.

RA MEDLINE=82046817; PubMed=6794640;

RA Metz-Boutigue M.-H., Mazurier J., Jolles J., Spik G., Montreuil J.,
Jolles P.;

RT "The present state of the human lactoferrin sequence. Study and
alignment of the cyanogen bromide fragments and characterization of
N- and C-terminal domains.";

RL Biochim. Biophys. Acta 670:243-254(1981).

RN [11]

RP SEQUENCE OF 609-711.

RA MEDLINE=82262043; PubMed=7049727;

RA Metz-Boutigue M.-H., Jolles J., Mazurier J., Spik G., Montreuil J.,
Jolles P.;

RT "An 88 amino acid long C-terminal sequence of human
lactoferrin.";

RL FEBS Lett. 142:107-110(1982).

RN [12]

RP SEQUENCE OF 436-711 FROM N.A.

RA MEDLINE=88001031; PubMed=3477300;

RA Rado T.A., Wei X., Benz E.J. Jr.;

RT "Isolation of lactoferrin cDNA from a human myeloid library and
expression of mRNA during normal and leukemic myelopoiesis.";

RL Blood 70:989-993(1987).

RN [13]

RP SEQUENCE OF 237-711 FROM N.A.

RA McCombie W.R., Wilson R., Chen E., Gibbs R., Zhu L., Johnson D.,
Nhan M., Parnell L., Dedhia N., Ansari A., Marlis E., Schutz K.,
Gao L., la Bastide M., Kaplan M., Greco T., Touchman J., Muzny D.,
Chen C.N., Evans C., Fitzgerald M., See L.H., Tang M., Porcel B.M.,
Dragan V., Giacalone J., Pae A., Powell E., Solinsky K.A., Desilva U.,
Diaz-Perez S., Zhou X., Yu Y., Watanabe M., Doggett N., Garcia D.,
Sagripanti J.L.;

RT Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

RN [14]

RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND REVISIONS.

RA MEDLINE=90064528; PubMed=2585506;

RA Anderson B.F., Baker H.M., Norris G.E., Rice D.W., Baker E.N.;

RT "Structure of human lactoferrin: crystallographic structure analysis
and refinement at 2.8-A resolution.";

RL J. Mol. Biol. 209:711-734(1989).

RN [15]

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).

RA Haridas M., Anderson B.F., Baker E.N.;

RT "Structure of human diferric lactoferrin refined at 2.2-A
resolution.";

RL Acta Crystallogr. D 51:629-646(1995).

RN [16]

RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 20-353.

RA MEDLINE=97156796; PubMed=9003186;

RA Nicholson H., Anderson B.F., Bland T., Shewry S.C., Tweedie J.N.,
Baker E.N.;

RT "Mutagenesis of the histidine ligand in human lactoferrin: iron
binding properties and crystal structure of the histidine-
253--methionine mutant.";

RL Biochemistry 36:341-346(1997).

RN [17]

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).

RA MEDLINE=99190892; PubMed=10089347;

RA Sun X.L., Baker H.M., Shewry S.C., Jameson G.B., Baker E.N.;

RT "Structure of recombinant human lactoferrin expressed in Aspergillus
awamori.";

```

RL Acta Crystallogr. D 55:403-407(1999).
RN [18]
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=99192677; PubMed=10089508;
RX Jameson G.B., Anderson B.F., Norris G.E., Thomas D.H., Baker E.N.;
RT "Structure of human apolactoferrin at 2.0-A resolution. Refinement
RT and analysis of ligand-induced conformational change.";
RX Acta Crystallogr. D 54:1319-1335(1998).
RN [19]
RP CHARACTERIZATION OF LACTOFERROXINS.
RX MEDLINE=91166929; PubMed=1369293;
RX Tani F., Ito K., Chiba H., Yoshikawa M.;
RT "Isolation and characterization of oploid antagonist peptides derived
RT from human lactoferrin.";
RX Agric. Biol. Chem. 54:1803-1810(1990).
RN [20]
RP VARIANTS THR-30 AND ARG-48.
RX MEDLINE=99091914; PubMed=9873069;
RX Kintworth G.K., Sommer J.R., O'Brien G., Han L., Ahmed M.N.,
RX Qumseyeh M.B., Lin P.-Y., Baeti S., Reddy M.K., Kanai A., Hotta Y.,
RX Sugai J., Kumaraswamykavel G., Munier P., Schorderet D.F.,
RX El Mariri L., Iwata F., Kaiser-Kupfer M., Nagata M., Nakayasu K.,
RX Hejtmancik J.F., Teng C.T.;
RT "Familial subepithelial corneal amyloidosis (gelatinous drop-like
RT corneal dystrophy): exclusion of linkage to lactoferrin gene.";
RX Mol. Vision 4:31-33(1998).
RN [21]
RP FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
RP CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
RP OF AN ANION. USUALLY BICARBONATE.
CC -1- FUNCTION: LACTOFERROXINS A, B AND C HAVE OPIOID ANTAGONIST
CC ACTIVITY. LACTOFERROXIN A SHOWS PREFERENCE FOR MU-RECEPTOR, WHILE
CC LACTOFERROXIN B AND C HAVE SOMEWHAT HIGHER DEGREES OF PREFERENCE
CC FOR KAPPA-RECEPTORS THAN FOR MU-RECEPTORS.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; X53961; CA37914.1; -
CC DR EMBL; U07643; AAB60324.1; -
CC DR EMBL; M93150; AAA56159.1; -
CC DR EMBL; M83202; AAA58511.1; -
CC DR EMBL; M83205; AAA58561.1; -

```

```

Query Match 61.0%; Score 100; DB 1; Length 711;
Best Local Similarity 57.1%; Pred. No. 4.2e-07;
Matches 16; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
QY 1 PEMSKCYQWQRMKXGAPITCIRTS 28
DB 34 PEATKCFQWQRMKXGAPVSCIRKDS 61

```

```

RESULT 6
TRFL_MOUSE STANDARD; PRT; 707 AA.
AC P08071; P70690; O61799; O922P2;
AC 01-AUG-1988 (Rel. 08, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lactotransferrin precursor (lactoferrin).
GN LTF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=uterus;
RX MEDLINE=87280033; PubMed=3611056;
RX Pentecost B.T., Teng C.T.;
RT "Lactotransferrin is the major estrogen inducible protein of mouse
RT uterine secretions.";
RX J. Biol. Chem. 262:10134-10139(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=uterus;
RX Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RX Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RX Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RX Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RX Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RX Stapleton M., Soares M.B., Bonaldo M.F., Caeavant T.L., Scheetz T.E.,
RX Brownstein M.J., Udell T.B., Toshiyuki S., Canninci P., Prange C.,
RX Raha S.S., Lottellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RX Boeak S.A., McEwan P.J., McKernan K.U., Malek J.A., Gunaratne P.H.,
RX Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Holys S.W.,
RX Villalón D.K., Muzny D.M., Sodergren E.U., Lu X., Gibbs R.A.,
RX Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RX Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RX Rodriguez A.C., Touchman J.W., Green E.D., Dickson M.C.,
RX Butterfield Y.S.N., Krzywicki M.T., Skalska U., Smallie D.E.,
RX Schercher A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 1-14 FROM N.A.
RX MEDLINE=92042099; PubMed=1939212;
RX Liu Y., Teng C.T.;
RT "Characterization of estrogen-responsive mouse lactoferrin promoter.";
RX J. Biol. Chem. 266:21880-21885(1991).
CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; U03298; AAA40525.1; -
CC DR EMBL; D88510; BAA13633.1; -
CC DR EMBL; BC006904; AAH06904.1; -
CC DR EMBL; M74778; AAA39427.1; -
CC DR HSSP; P02788; ICB6.
CC MGD; MGI:96837; Ltf.
CC InterPro; IPR001156; Transferrin.
CC Pfam; PF00405; Transferrin; 2.
CC PRINTS; PM00422; TRANSFERRIN.
CC SMART; SM00094; TR_FER; 2.
CC PROSITE; PS00205; TRANSFERRIN_1; 1.
CC PROSITE; PS00206; TRANSFERRIN_2; 2.
CC PROSITE; PS00207; TRANSFERRIN_3; 2.
CC Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;

```

KM Signal. 1 19 BY SIMILARITY.
 FT SIGNAL. 20 707 LACTOTRANSFERRIN.
 RT CHAIN 20 357 1.
 FT REPEAT 358 707 2.
 FT REPEAT 358 707 2.
 FT DISULFID 27 63 BY SIMILARITY.
 FT DISULFID 27 63 BY SIMILARITY.
 FT DISULFID 37 54 BY SIMILARITY.
 FT DISULFID 133 216 BY SIMILARITY.
 FT DISULFID 175 191 BY SIMILARITY.
 FT DISULFID 188 199 BY SIMILARITY.
 FT DISULFID 199 263 BY SIMILARITY.
 FT DISULFID 249 398 BY SIMILARITY.
 FT DISULFID 366 389 BY SIMILARITY.
 FT DISULFID 376 702 BY SIMILARITY.
 FT DISULFID 423 702 BY SIMILARITY.
 FT DISULFID 443 665 BY SIMILARITY.
 FT DISULFID 475 550 BY SIMILARITY.
 FT DISULFID 499 523 BY SIMILARITY.
 FT DISULFID 509 533 BY SIMILARITY.
 FT DISULFID 520 533 BY SIMILARITY.
 FT DISULFID 591 605 BY SIMILARITY.
 FT DISULFID 643 648 BY SIMILARITY.
 FT METAL 78 78 IRON 1 (BY SIMILARITY).
 FT METAL 110 110 IRON 1 (BY SIMILARITY).
 FT METAL 210 210 IRON 1 (BY SIMILARITY).
 FT METAL 271 271 IRON 1 (BY SIMILARITY).
 FT METAL 413 413 IRON 2 (BY SIMILARITY).
 FT METAL 451 451 IRON 2 (BY SIMILARITY).
 FT METAL 544 544 IRON 2 (BY SIMILARITY).
 FT METAL 613 613 IRON 2 (BY SIMILARITY).
 FT BINDING 139 139 ANION (POTENTIAL).
 FT BINDING 481 481 ANION (POTENTIAL).
 FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 494 494 MR -> ICG (IN REF. 1).
 FT CONFLICT 1 2 R -> Q (IN REF. 2).
 FT CONFLICT 25 25 M -> L (IN REF. 2).
 FT CONFLICT 82 82 S -> T (IN REF. 2).
 FT CONFLICT 359 359 A -> D (IN REF. 1).
 FT CONFLICT 382 382 E -> G (IN REF. 2).
 FT CONFLICT 449 449 L -> V (IN REF. 1).
 FT CONFLICT 629 629
 SQ SEQUENCE 707 AA; 77865 MW; F26AE0340A4C19A8 CRC64;

Query Match 51.2%; Score 84; DB 1; Length 707;
 Best Local Similarity 44.4%; Pred. No. 9.8e-05;
 Matches 12; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

OY 2 EMSKCYOMRRMRKLGAPISITCIRTS 28
 DB 33 EEKCLRMQNMRRKVGCPPLSCVKSS 59

RESULT 7
 TRFL HORSE STANDARD; PRT; 695 AA.
 ID 077811;
 AC 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Lactotransferrin precursor (Lactoferrin) (Fragment).
 GN LTF.
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Euteria; Perissodactyla; Equidae; Equus.
 OX NCBI_TaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Paramesivam M., Srinivasan A., Yadav M.P., Singh T.P.;
 RT "CDNA sequence of mare lactoferrin."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBS databases.
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RC TISSUE=Milk;
 RX MEDLINE=993296631; PubMed=1036507;

RA Sharma A.K., Paramesivam M., Srinivasan A., Yadav M.P., Singh T.P.;
 RT "Three-dimensional structure of mare dimeric lactoferrin at 2.6-A
 resolution."
 RL J. Mol. Biol. 289:303-317(1999).
 CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION, USUALLY BICARBONATE.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ010930; CA009407.1; -.
 DR PDB; 1BX; 02-DEC-98.
 DR PDB; 1B7U; 02-FEB-99.
 DR PDB; 1B7Z; 02-FEB-99.
 DR PDB; 1P98; 10-FEB-01.
 DR PDB; 1I6B; 13-FEB-02.
 DR PDB; 1QJN; 14-JAN-00.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 1.
 DR Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 KW Signal; Jd-structure.
 KM SIGNAL 1
 FT NON_TER 1
 FT SIGNAL <1 695
 FT CHAIN 7 695
 FT REPEAT 7 350
 FT REPEAT 351 695
 FT DISULFID 15 51
 FT DISULFID 25 42
 FT DISULFID 121 204
 FT DISULFID 163 179
 FT DISULFID 166 189
 FT DISULFID 176 187
 FT DISULFID 237 251
 FT DISULFID 354 386
 FT DISULFID 364 377
 FT DISULFID 411 690
 FT DISULFID 431 653
 FT DISULFID 463 538
 FT DISULFID 487 681
 FT DISULFID 497 511
 FT DISULFID 508 521
 FT DISULFID 579 593
 FT DISULFID 631 636
 FT METAL 66
 FT METAL 66
 FT METAL 98
 FT METAL 98
 FT METAL 198
 FT METAL 259
 FT METAL 401
 FT METAL 439
 FT METAL 532
 FT METAL 532
 FT METAL 601
 FT METAL 601
 FT BINDING 127 127
 FT BINDING 469 469
 FT CARBOHYD 143 143
 FT CARBOHYD 287 287
 FT CARBOHYD 482 482
 FT STRAND 12 16
 FT HELIX 19 34

IRON 1 (BY SIMILARITY).
 IRON 1 (BY SIMILARITY).
 IRON 1 (BY SIMILARITY).
 IRON 1 (BY SIMILARITY).
 IRON 2 (BY SIMILARITY).
 IRON 2 (BY SIMILARITY).
 IRON 2 (BY SIMILARITY).
 ANION (BY SIMILARITY).
 ANION (BY SIMILARITY).
 ANION (BY SIMILARITY).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 LACTOTRANSFERRIN.
 1.
 2.


```

FT TURN 35 36
FT STRAND 40 44
FT HELIX 48 56
FT TURN 57 58
FT STRAND 62 65
FT HELIX 67 74
FT TURN 76 78
FT STRAND 80 88
FT STRAND 95 95
FT STRAND 97 105
FT TURN 106 107
FT HELIX 112 114
FT TURN 116 117
FT STRAND 119 122
FT TURN 125 126
FT TURN 128 131
FT HELIX 132 137
FT HELIX 139 142
FT TURN 147 148
FT HELIX 151 158
FT STRAND 161 163
FT TURN 165 166
FT TURN 169 171
FT HELIX 173 175
FT TURN 176 178
FT TURN 183 187
FT TURN 191 192
FT TURN 194 195
FT HELIX 197 206
FT TURN 207 208
FT STRAND 212 216
FT HELIX 219 223
FT HELIX 227 230
FT TURN 231 232
FT STRAND 233 237
FT TURN 238 240
FT STRAND 241 244
FT HELIX 245 247
FT TURN 248 250
FT STRAND 254 257
FT STRAND 260 264
FT HELIX 270 284
FT TURN 286 287
FT TURN 294 295
FT TURN 308 309
FT STRAND 312 315
FT HELIX 322 326
FT HELIX 328 336
FT TURN 337 338
FT HELIX 341 349
FT STRAND 351 356
FT HELIX 358 370
FT TURN 371 372
FT STRAND 375 380
FT HELIX 383 392
FT TURN 393 393
FT STRAND 397 400
FT HELIX 402 410
FT TURN 411 412
FT STRAND 414 421
FT HELIX 429 433
FT STRAND 439 446
FT HELIX 454 456
FT TURN 458 459
FT STRAND 462 464
FT TURN 467 468
FT TURN 470 473
FT HELIX 474 484
FT HELIX 489 491
FT TURN 492 492
FT STRAND 496 497
FT TURN 499 500
FT TURN 503 504

```

```

FT TURN 506 507
FT STRAND 508 508
FT TURN 515 516
FT TURN 520 521
FT TURN 525 526
FT TURN 528 529
FT HELIX 531 540
FT TURN 541 542
FT STRAND 546 550
FT HELIX 551 555
FT TURN 557 557
FT TURN 559 560
FT HELIX 565 568
FT TURN 569 569
FT TURN 572 573
FT STRAND 575 578
FT TURN 580 581
FT STRAND 584 586
FT HELIX 587 592
FT STRAND 596 598
FT TURN 602 605
FT TURN 607 609
FT HELIX 610 624
FT TURN 626 627
FT TURN 629 630
FT HELIX 631 634
FT TURN 637 638
FT TURN 641 642
FT TURN 649 650
FT STRAND 651 655
FT TURN 658 659
FT HELIX 663 667
FT HELIX 669 679
FT TURN 680 681

```

Query Match 50.6%; Score 83; DB 1; Length 695;
 Best Local Similarity 50.0%; Pred. No. 0.0013;
 Matches 14; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

Qy 2 EMSKCYOMRMRKLGAPSTTCIRRTSA 29
 Db 21 EAACCAKFGAMKVGPSVSCIRKTS 48

```

RESULT 8
TRFL_PIG ID TRFL_PIG STANDARD; PRT; 704 AA.
AC P14632; Q29557;
DT 01-APR-1990 (Rel. 14, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactoferrin precursor (Lactoferrin).
GN LTF.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OC NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92379101; PubMed=1511016;
RA Lyndon J.P., O'Malley B.R., Saucedo O., Lee T., Heaton D.R.,
RA Conneely O.M.;
RT "Nucleotide and primary amino acid sequence of porcine lactoferrin.";
RL Biochim. Biophys. Acta 1132:97-99(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92367939; PubMed=1503259;
RA Alexander L.J., Levine W.B., Teng C.T., Beattie C.W.;
RT "Cloning and sequencing of the porcine lactoferrin cDNA.";
RL Anim. Genet. 23:251-256(1992).
RN [3]
RP SEQUENCE OF 20-49.
RX MEDLINE=90105538; PubMed=2605266;

```


DT 28-FEB-2003 (Rel. 41, last sequence update)
 DT 15-SEP-2003 (Rel. 42, last annotation update)
 DE Putative mannose-3-phosphoglycerate phosphatase (EC 3.1.3.70) (MGP)
 GN YEDP OR STM1986 OR STY2193 OR T0892.
 OS *Salmonella typhimurium*, and
 CC *Salmonella typhi*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; *Salmonella*.
 OK NCBI_TaxID=602, 601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Clifton J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dame M., Du P., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvany E.,
 RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 LT2.";
 RL Nature 413:852-856(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. typhi; STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Richard D., Main J.,
 RA Churcher S., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaiha M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin P., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. typhi; STRAIN=TY2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodymani V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- CATALYTIC ACTIVITY: 2(alpha-D-mannosyl)-3-phosphoglycerate + H(2)O
 CC = 2(alpha-D-mannosyl)-D-glycerate + phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE HAD SUPERFAMILY. MGP FAMILY.
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL, AB008786; AAL20896.1; -
 CC EMBL, AL627272; CAD05733.1; -
 CC EMBL, AE016837; AAO68570.1; -
 CC STyGene, SG727272; yedp.
 CC HAMAP, MF_00617; -; 1.
 CC InterPro, IPR006379; HAD_SF_TTB.
 CC InterPro, IPR006381; HAD_SF_Yedp.
 CC InterPro, IPR005834; Hydrolyase.
 CC Pfam, PF00702; Hydrolyase; 1.
 CC TIGRFAMs, TIGR01486; HAD-SF-TTB-Yedp; 1.
 CC TIGRFAMs, TIGR01484; HAD-SF-TTB; 1.
 CC Hypothetical protein; Hydrolyase; Complete proteome.
 CC SEQUENCE 271 AA; 30925 MW; 46866C2E36B7008E CRC64;
 RP
 RP Query Match 35.4%; Score 58; DB 1; Length 271;

Best Local Similarity 35.7%; Pred. No. 0.25;
 Matches 10; Conservative, 4; Mismatches 14; Indels 0; Gaps 0;
 QY 2 EWSKCYQWRMRKLGAPSIITCRRTSA 29
 DB 25 EWCPAAPMLTRHESGVPVILCSKTA 52
 RESULT 13
 TRFM_HUMAN
 ID TRFM_HUMAN STANDARD; PRT; 738 AA.
 AC P08582; Q9B0E2;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 15-SEP-2003 (Rel. 42, last annotation update)
 DE Melanotransferrin precursor (Melanoma-associated antigen p97) (CD228
 DE antigen).
 GN MF12 OR MAP97.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OK NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Melanoma;
 RX MEDLINE=86149285; PubMed=2419904;
 RA Rose T.M., Plozman G.D., Teplow D.B., Dreyer W.J., Hellstrom K.E.,
 RA Brown J.P.;
 RT "Primary structure of the human melanoma-associated antigen p97
 RT (melanotransferrin) deduced from the mRNA sequence.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:1261-1265(1986).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Skin, and Uterus;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bat N.K.,
 RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marisina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stadelman M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carinai P., Prange C.,
 RA Rahn S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Boeck S.A., McEwan P.J., McKernan K.J., Malek J.A., Gannatare P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schercher A., Schein J.B., Jones S.J.M., Marra M.A.; full-length
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP GPI-ANCHOR.
 RX MEDLINE=94132080; PubMed=8300636;
 RA Food M.R., Kochenberger S., Gabathuler R., Haid I.D., Reid G.,
 RA Jufferies W.A.;
 RT "Transport and expression in human melanomas of a transferrin-like
 RT glycosylphosphatidylinositol-anchored protein.";
 RL J. Biol. Chem. 269:3034-3040(1994).
 RN [4]
 RP FUNCTION.
 RX MEDLINE=96016189; PubMed=7556058;
 RA Kennard M.L., Richardson D.R., Gabathuler R., Ponka P.,
 RA Jufferies W.A.;
 RT "A novel iron uptake mechanism mediated by GPI-anchored human p97.";
 RL EMBO J. 14:4178-4186(1995).
 RN [5]
 RP IRON-BINDING.
 RX MEDLINE=92183868; PubMed=1544447;

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U25051; AAA67686.1; -
 DR EMBL; U25046; AAA67686.1; JOINED.
 DR EMBL; U25047; AAA67686.1; JOINED.
 DR EMBL; U25048; AAA67686.1; JOINED.
 DR EMBL; U25049; AAA67686.1; JOINED.
 DR EMBL; U25050; AAA67686.1; JOINED.
 DR MGD; MGI:104535; PENT. 1.
 DR Pfam; PF04191; PENT. 1.
 KW Phospholipid biosynthesis; Transferase; Methyltransferase;
 KW Transmembrane; Mitochondrion; Endoplasmic reticulum.
 FT INIT_MET 0
 FT TRANSMEM 12 32 POTENTIAL.
 FT TRANSMEM 45 65 POTENTIAL.
 FT TRANSMEM 90 110 POTENTIAL.
 FT TRANSMEM 157 177 POTENTIAL.
 SQ SEQUENCE 198 AA; 22385 MW; 5828975A5173D72 CRC64;

Query Match 31.7%; Score 52; DB 1; Length 198;
 Best Local Similarity 36.0%; Pred. No. 1.4;
 Matches 9; Conservative 6; Mismatches 6; Indels 4; Gaps 1;

QY 3 WSKCYQWQRMRKRL---GAPSIYC 23
 Db 28 WNVVAPRWQRTRKLSRAFGSPHLAC 52

RESULT 15
 MGP_ECOL6
 ID MGP_ECOL6 STANDARD; PRT; 271 AA.
 AC P59286;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative mannosyl-3-phosphoglycerate phosphatase (EC 3.1.3.70) (MGP).
 GN YEDP OR C2373.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Raske D., Buckland E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Domeneberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 RL
 CC -1- CATALYTIC ACTIVITY: 2(alpha-D-mannosyl)-3-phosphoglycerate + H(2)O
 CC = 2(alpha-D-mannosyl)-D-glycerate + phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE HAD SUPERFAMILY. MGP FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE016762; AAN80832.1; -
 DR HAMAP; MF_00617; -; 1.

DR Pfam; PF00702; Hydrolase; 1.
 DR TIGRFAMs; TIGR01486; HAD-SF-TIB-YedP; 1.
 DR TIGRFAMs; TIGR01484; HAD-SF-TIB; 1.
 KW Hypothetical protein; Hydrolase; Complete proteome.
 SQ SEQUENCE 271 AA; 30515 MW; E6EB78B391470B2 CRC64;

Query Match 31.7%; Score 52; DB 1; Length 271;
 Best Local Similarity 32.1%; Pred. No. 2;
 Matches 9; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY 2 WSKCYQWQRMRKRLGAPSIYTCIRRTSA 29
 Db 25 DWGPAPWLSRLHEANIPLYLCSKRTSA 52

Search completed: July 30, 2003, 16:24:31
 Job time : 10.494 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using bw model

Run on: July 30, 2003, 16:16:55 ; Search time is 4702 seconds

(without alignment)
150,994 Million cell updates/sec

Title: US-09-787-070-6

Perfect score: 164

Sequence: 1 PEMSKCYOMORMRKLGAPSTICIRRTSA 29

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

1: PIR 76:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	163	99.4	708	2 JC2323	lactoferrin - goat
2	122	74.4	708	1 TFRBL	lactotransferrin p
3	100	61.0	711	1 TFRHU	lactotransferrin p
4	91	55.5	33	2 S52107	lactoferrin - sheep
5	84	51.2	707	1 A45438	lactoferrin precursor
6	66	40.2	703	2 A45543	lactoferrin precursor
7	58.5	35.7	704	2 147228	carbonic anhydrase
8	58	35.4	271	2 B85813	hypothetical prote
9	58	35.4	271	2 E90965	hypothetical prote
10	58	35.4	271	2 G64959	conserved hypotet
11	58	35.4	271	2 AB0754	melanotransferrin
12	57.5	35.1	738	1 TFRHM	phosphatidylethano
13	51	31.1	199	2 A47353	probable monooxyge
14	51	31.1	489	2 C70655	transferrin precu
15	50	30.5	706	2 S33761	B. subtilis yabp p
16	49	29.9	405	2 AB1461	B. subtilis yabp p
17	49	29.9	408	2 AC1098	tryptophan halogen
18	49	29.9	507	2 G87466	hypothetical prote
19	49	29.9	1721	2 T21214	histidyl tRNA synt
20	48.5	29.6	71	2 AH0033	histidyl tRNA 11g
21	48	29.3	430	2 H72052	F2H15.8 protein -
22	48	29.3	430	2 H72052	noar protein - Par
23	47.5	29.0	423	2 F86313	ecarotenom (CER3)
24	47	28.7	98	2 S39410	hypothetical prote
25	47	28.7	528	2 T21634	hypothetical prote
26	47	28.7	795	2 T48252	hypothetical prote
27	47	28.7	822	2 AB2507	histocytol 1,4,5-tri
28	47	28.7	2671	2 A49873	hypothetical prote
29	46.5	28.4	121	2 AH3147	hypothetical prote

30	46.5	28.4	694	1 TFRBP	transferrin precu
31	46	28.0	413	2 AH0780	probable oxidoredu
32	46	28.0	469	2 T26463	hypothetical prote
33	46	28.0	488	2 S36570	E2 protein - human
34	46	28.0	622	2 S61692	probable membrane
35	46	28.0	647	2 E82126	DNA topoisomerase
36	46	28.0	1082	2 T50650	AP-3 complex beta3
37	45.5	27.7	384	2 E69050	glcNAc-phosphatidy
38	45.5	27.7	550	2 T47158	hypothetical prote
39	45.5	27.7	620	2 T27008	hypothetical prote
40	45.5	27.7	698	1 TFRHP	transferrin precu
41	45	27.4	200	1 A42725	nitrite hydratase
42	45	27.4	284	2 S58650	hypothetical prote
43	45	27.4	310	2 C6218	transcription init
44	45	27.4	368	2 T55961	MHC class I histoc
45	45	27.4	395	2 T31578	hypothetical prote

ALIGNMENTS

RESULT 1

JC2323

C/Species: Capra aegagrus hircus (domestic goat)

C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 07-May-1999

C/Accession: JC2323

R/Le Provost, F.; Nocart, M.; Guerin, G.; Martin, P.

Biochem. Biophys. Res. Commun. 203, 1324-1332, 1994

A/Title: Characterization of the goat lactoferrin cDNA: assignment of the relevant locu

A/Reference number: JC2323; MUID:94380047; PMID:8093048

A/Accession: JC2323

A/Molecule type: mRNA

A/Residues: 1-708 <UEP>

C/Superfamily: transferrin; transferrin repeat homology

C/Keywords: duplication; glycoprotein

F/359-696/Domain: transferrin repeat homology <TR32>

F/252,300,387,495,564/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match

Best Local Similarity 99.4%; Score 163; DB 2; Length 708;

Matches 28; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PEMSKCYOMORMRKLGAPSTICIRRTSA 29

DB 33 PEMSKCYOMORMRKLGAPSTICIRRTSA 61

RESULT 2

TFRBL

lactotransferrin precursor - bovine

N/Alternate names: lactoferrin

C/Species: Bos primigenius taurus (cattle)

C/Date: 31-Mar-1992 #sequence_revision 21-Nov-1997 #text_change 11-May-2000

C/Accession: I45919; S14674; S14110; S18517; U05055; S13097; S18518; S13881; E0148; S221

R/Tsang, T.C.; Burns, D.K.; Wang, F.; Pan, Y.

FASEB J. 6, 233, 1991

A/Title: Cloning of a 80-kD advanced glycosylation end product (AGE) binding protein from

A/Reference number: I45919

A/Accession: I45919

A/Status: translated from GB/EWBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-708 <TSA>

A/Cross-references: GB:I08604; NID:g163269; PIDN:AAA30609.1; PID:g163270

R/Pierce, A.

submitted to the EMBL Data Library, November 1990

A/Reference number: S14674

A/Accession: S14674

A/Molecule type: mRNA

A/Residues: 1-144, 'V', 146-163, 'PP', 166-339, 'A', 341-438, 'Y', 440-513, 'R', 515-708 <PI>

A/Cross-references: EMBL:X57084; NID:g505; PIN:CAA40366.1; PID:g506

R/Pierce, A.; Colavizza, D.; Benaisse, M.; Maes, P.; Tartar, A.; Montreuil, J.; Spik, G.

Eur. J. Biochem. 196, 177-184, 1991


```

A>Status: not compared with conceptual translation
A:Molecule type: mRNA
R:Residues: 3-701,'SMKPVN' <PAN>
A:Experimental source: normal breast tissue
R:Meiz-Boutigue, M.H.; Jolles, V.; Mazurier, J.; Schoentgen, F.; Legrand, D.; Spik, G.;
Bur, J. Biochem. 145, 659-666, 1984
A>Title: Human lactotransferrin: amino acid sequence and structural comparisons with other
A:Reference number: A31000; PMID:85076667; PMID:6510420
A:Accession: A31000
A:Molecule type: protein
A:Residues: 20-140,142-169,171-203,'L',205,'K',207-208,'K',210-385,'Q',387-391,'W',393-406
A>Note: This is the final paper in a series
R:Houen, G.; Hoegdall, B.V.; Barkholt, V.; Norrskov, L.
Eur. J. Biochem. 241, 303-308, 1996
A>Title: Lactoferrin: similarity to diamine oxidase and purification by aminohexyl affinity
A:Reference number: S74119; PMID:97054624; PMID:8898921
A:Accession: S74119
A:Molecule type: protein
A:Residues: 'G',23-24,'R',26-27,'XX',30-32 <HOU>
A:Experimental source: neutrophil granulocytes
C:Genetics:
A:Gene: GDB:LTF
A:Cross-references: GDB:119368; OMIM:150210
A:Map position: 3q21-3q23
C:Superfamily: transferrin; transferrin repeat homology
C:Keywords: duplication; glycoprotein; iron binding; milk
F:1-19/Domains: signal sequence #status predicted <SIG>
F:20-711/Product: lactotransferrin #status experimental <MAT>
F:21-356/Domains: transferrin repeat homology <TRH1>
F:360-689/Domains: transferrin repeat homology <TRH2>
F:129-65,39-56,133-218,177-193,190-201,251-265,503-697,595-609/Disulfide bonds: #status e
F:157,458/Binding sites: carbohydrate (Asn) (covalent) #status experimental
F:368-400,378-391,425-706,447-669,479-554,513-527,524-537,647-652/Disulfide bonds: #statu

```

C:\Species: Mus musculus (house mouse) C:\Species: Mus musculus (house mouse)
C:\Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:\Accession: A28438, A41205
R:\pentecost, B.T.; Teng, C.T.
U. Biol. Chem. 262, 10134-10139, 1987
A:\Title: Lactotransferrin is the major estrogen-inducible protein of mouse uterine secretory granules
A:\Reference number: A92556; PMID:87280033; PMID:3611056
A:\Accession: A28438
A:\Molecule type: mRNA
A:\Residues: 3-707 <PEN>
A:\Cross-references: EMBL:J03298
R:\Lin, Y.; Teng, C.T.
U. Biol. Chem. 266, 21880-21885, 1991
A:\Title: Characterization of estrogen-responsive mouse lactoferrin promoter.
A:\Reference number: A41205; PMID:92042099; PMID:1939212
A:\Accession: A41205
A:\Molecule type: DNA
A:\Residues: 1-15 <LTU>
A:\Cross-references: GB:M74778
C:\Superfamily: transferrin; transferrin repeat homology
C:\Keywords: duplication; glycoprotein
F:\1-19/Domain: signal sequence #status predicted <SIG>
F:\20-707/Product: lactotransferrin #status predicted <MAT>
F:\358-695/Domain: transferrin repeat homology <TRH>
F:\494/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 51.2% Score 84; DB 1; Length 707;
Best Local Similarity 44.4%; Pred. NO. 0.00027;
Matches 12; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Qy 2 EWKCYOWRRNRKLGAFTICRTS 28
| | | | | : : : :
Db 33 EEKCI.RWQNEKRKYGGPPLSCVKSS 59

RESULT 6
A45543
lactoferrin precursor - pig
C:\Species: Sus scrofa domestica (domestic pig)
C:\Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 04-Mar-2000
C:\Accession: A45543; S24173
R:\Alexander, L.D.; Levine, W.B.; Teng, C.T.; Beattie, C.W.
Anim. Genet. 23, 251-256, 1992
A:\Title: Cloning and sequencing of the porcine lactoferrin cDNA.
A:\Reference number: A45543; PMID:92367939; PMID:1503259
A:\Accession: A45543
A:\Status: preliminary
A:\Molecule type: mRNA
A:\Residues: 1-703 <ALB>
A:\Experimental source: mammary gland
A:\Note: sequence extracted from NCBI backbone (NCBIN:111151, NCBIP:111153)
R:\Lydon, T.P.; O'Malley, B.R.; Saucedo, O.; Lee, T.; Haddon, D.R.; Connely, O.M.
Biochim. Biophys. Acta 1132, 97-99, 1992
A:\Title: Nucleotide and primary amino acid sequence of porcine lactoferrin.
A:\Reference number: S24173; PMID:92379101; PMID:1511016
A:\Accession: S24173
A:\Molecule type: mRNA
A:\Residues: 1-11, 'W', 'I', '52-84', 'G', '86-120', 'V', '121-130', 'I', '132-282', 'S', '284-571', 'Q', 'E', '572-691'
A:\Cross-references: EMBL:M92089, NID:g164613; PIDN:AAA31102.1; PID:g164614
A:\Experimental source: mammary gland
C:\Superfamily: transferrin; transferrin repeat homology
C:\Keywords: duplication; glycoprotein; iron; iron binding; metalloprotein; milk
F:\1-19/Domain: signal sequence #status predicted <SIG>
F:\20-703/Product: lactoferrin #status predicted <MAT>
F:\20-350/Domain: transferrin repeat homology <TRH>
F:\36-48/Region: antimicrobial
F:\354-691/Domain: transferrin repeat homology <TRH>
F:\28-62, 38-53, 129-212, 171-187, 184-195, 245-259, 362-394
F:\77, 107, 206, 267/Binding site: iron (Asp, Tyr, Tyr, His) #status predicted
F:\115/Binding site: carbonate (Arg) #status predicted
F:\409, 447, 540, 609/Binding site: iron (Asp, Tyr, Tyr, His) #status predicted
F:\477/Binding site: carbonate (Arg) #status predicted
F:\450/Binding site: carbohydrate (Asn) (covalent) #status predicted

C/Species: *Escherichia coli*
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 14-Dec-2001
C/Accession: E90965
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
R/Gasawara, N.; Yessunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic
Reference number: A996293, MUID:21156231, PMID:11258796

Query Match	35.4%	Score 58	DB 2	Length 271
Best Local Similarity	35.7%	Pred. No. 0.71		
Matches	10	Conservative	4	Mismatches 14
				Indels 0
				Gaps 0
QY	2	EWSCKYQWQRRKRGKAPSTICIRRTSA	29	
	:	:		
Db	25	DMQAPAPMISLRREANVEVILCSKTSIA	52	

RESULT 10
 G64959
 hypothetical protein b155 - *Escherichia coli* (strain K-12)
 C|Species: *Escherichia coli*
 C|Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #ext_change 01-Mar-2002
 C|Accession: G64959
 R|Blatner, P.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co-
 A.; Rose, D.U.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A|Title: The complete genome sequence of *Escherichia coli* K-12.
 A|Reference number: A64720; MUID:9742617; EMD:9278503

A/Residues: 1-271 >BLAT
A/Accession: G3:AE000287; GB:U00096; NID:g1788257; PID:g1788265
A/Cross-references: GB:AE000287; GB:U00096; NID:g1788257; PID:g1788265
A/Experimental source: strain K-12, substrain MG1655
C/Superfamily: Pyrococcus horikoshii hypothetical protein PH0926
Query Match 35.4%; Score 58; DB 2; Length 271;
Best Local Similarity 35.7%; Pred. NO. 0.71;
Matches 10; Conservative 4; Mismatches 14; Indels 0; Gaps 0;
2 EMSKCYOWGRNRKLGAPSIICIRRTSA 29

AB0754 conserved hypothetical protein STY2193 [imported] - *Salmonella enterica* subsp. *enterica*
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typh
A>Note: this species has also been called *Salmonella typhi*
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AB0754
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher
T.; S. Monteiro, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar
S.; Moule, S.; O'Garra, P.
A:Authors: PARRY, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* sero

A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* sero
A:Authors: Parr, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skellern, J.; Stevens, K
A:Accession: AB0754
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-271 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05733.1; PID:G16503226; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY2193
 C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0926
 Query Match 35.4%; Score 58; DB 2; Length 271;
 Best Local Similarity 35.7%; Pred. No. 0.71;
 Matches 10; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY 2 EMSKCQWQRRMRKLGAPSTICIRRTSA 29
 DB 25 EMQPADWLTREHSGVPYLCSKTA 52

RESULT 12
 TFHM
 melanotransferrin precursor - human
 N:Alternate names: melanoma-associated antigen gp95/p97
 C:Species: Homo sapiens (man)
 C>Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 28-Jan-2000
 C:Accession: A23814; A60925
 R:Rose, T.M.; Plowman, G.D.; Teplow, D.B.; Dreyer, W.J.; Hellstrom, K.E.; Brown, J.P.
 Proc. Natl. Acad. Sci. U.S.A. 83, 1261-1265, 1986
 A:Title: Primary structure of the human melanoma-associated antigen p97 (melanotransferrin)
 A:Reference number: A23814; MUID:86149285; PMID:2419904
 A:Accession: A23814
 A:Molecule type: mRNA
 A:Residues: 1-738 <RDS>
 A:Cross-references: EMBL:M12154; NID:G189515; PIDN:AAAS9992.1; PID:G189518
 A:Experimental source: melanoma
 R:Funakawa, K.S.; Funakawa, K.; Real, F.X.; Old, L.D.; Lloyd, K.O.
 J. Exp. Med. 169, 585-590, 1989
 A:Title: A unique antigenic epitope of human melanoma is carried on the common melanoma
 A:Reference number: A60925; MUID:8604252; PMID:2463331
 A:Accession: A60925
 A:Molecule type: Protein
 A:Residues: 20-25,'X',27-28,'X',30 <FUR>
 C:Comment: This protein is found predominantly in human melanomas and in certain fetal
 C:Comment: Seven disulfide bonds are predicted in each domain.
 C:Genetics:
 A:Gene: GDB:MFR2
 A:Cross-references: GDB:119387; OMIM:155750
 A:Map position: 3q28-3q29
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keywords: blocked carboxyl end; duplication; glycoprotein; lipoprotein; membrane prote
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:18-360/Domain: transferrin repeat homology <TRH1>
 F:20-710/Product: melanotransferrin #status predicted <WTF>
 F:361-709/Domain: transferrin repeat homology <TRH2>
 F:711-738/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F:38,135,515/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:710/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Set) (in mature form

Query Match 35.1%; Score 57.5; DB 1; Length 738;
 Best Local Similarity 43.3%; Pred. No. 2.2;
 Matches 13; Conservative 3; Mismatches 13; Indels 1; Gaps 1;

QY 1 PEMSKCQWQRRMRKLG-APSICIRRTSA 29
 DB 31 PEQHKCGMSAEAREAGTDPSTLCVGRGTA 60

RESULT 13
 A47353
 phosphatidylethanolamine N-methyltransferase (EC 2.1.1.17), 20K - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 08-May-1995 #sequence_revision 12-May-1995 #text_change 05-May-2000
 C:Accession: A47353
 R:Chui, Z.; Vance, J.E.; Chen, M.H.; Voelker, D.R.; Vance, D.E.
 J. Biol. Chem. 268, 1665-1669, 1993
 A:Title: Cloning and expression of a novel phosphatidylethanolamine N-methyltransferase.
 A:Reference number: A47353; MUID:9346418; PMID:8344945
 A:Accession: A47353

A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-199 <CUI>
 A:Cross-references: GB:U14441; NID:G310194; PIDN:AAA03154.1; PID:G310195
 C:Superfamily: phosphatidyl-N-methylethanolamine N-methyltransferase
 C:Keywords: membrane protein; methyltransferase; S-adenosylmethionine

Query Match 31.1%; Score 51; DB 2; Length 199;
 Best Local Similarity 36.0%; Pred. No. 5.6;
 Matches 9; Conservative 6; Mismatches 6; Indels 4; Gaps 1;

QY 3 WSKCYQWQRRMRKLG-APSITC 23
 DB 29 WNVVAREWQRRKRLSRAPSPYLAC 53

RESULT 14
 C70655
 Probable monooxygenase - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: C70655
 R:Coile, S.T.; Broese, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
 J. Mol. Biol. 281, 1-11, 1998
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: C70655
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-489 <COL>
 A:Cross-references: GB:283864; GB:AL123456; NID:G3261687; PIDN:CAE06212.1; PID:G301250;
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV3854c

Query Match 31.1%; Score 51; DB 2; Length 489;
 Best Local Similarity 57.1%; Pred. No. 13;
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 WSKCYQWQRRMRK 16
 DB 250 YSACQWQRRMRK 263

RESULT 15
 S33761
 transferrin precursor - horse
 N:Alternate names: growth-promoting factor
 C:Species: Equus caballus (domestic horse)
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
 C:Accession: S33761; S02145
 R:Carpenter, M.A.; Broad, T.E.
 Biochim. Biophys. Acta 1173, 230-232, 1993
 A:Title: The cDNA sequence of horse transferrin.
 A:Reference number: S33761; MUID:93277958; PMID:8504171
 A:Accession: S33761
 A:Molecule type: mRNA
 A:Residues: 1-706 <CAR>
 A:Cross-references: EMBL:M69020; NID:G164242; PIDN:AAA0958.1; PID:G164243
 A:Experimental source: liver; developmental stage adult
 R:Yoshinari, K.; Yuasa, K.; Iga, F.; Mitsuura, A.
 Biochim. Biophys. Acta 1010, 28-34, 1989
 A:Title: A growth-promoting factor for human myeloid leukemia cells from horse serum idc
 A:Reference number: S02145; MUID:89076697; PMID:2905248
 A:Accession: S02145
 A:Molecule type: Protein
 A:Residues: 20-35,'X',37,'X',39-40,'X',43-44 <YOS>
 C:Complex: monomer
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keywords: duplication; glycoprotein; iron transport; metal binding; plasma

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:09:50 ; Search time 57.1369 Seconds

(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-6

Perfect score: 164
Sequence: 1 PEMSKCYQMORRMKLGAPSTICIRRTSA 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*
1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	164	100.0	29	21	AA88341
2	164	100.0	708	20	AA880774
3	163	99.4	708	20	AA880775
4	136	82.9	25	21	AA883112
5	136	82.9	25	21	AA885013
6	122	74.4	689	21	AA49270
7	122	74.4	708	17	AAW09343
8	122	74.4	708	19	AAW57318
9	122	74.4	708	20	AAW86022

10	122	74.4	708	21	AA808183	Amino acid sequenc
11	122	74.4	708	22	AA664827	Chronic hepatitis
12	122	74.4	708	22	AAE02342	Bovine lactoferrin
13	120	73.2	32	14	AA84845	Lactoferrin-relate
14	120	73.2	32	15	AA848534	Lactoferrin-derive
15	120	73.2	32	15	AA857466	Lactoferrin derive
16	120	73.2	32	16	AA884703	Bovine lactoferrin
17	120	73.2	32	16	AA880268	Anti-parasitic lac
18	120	73.2	32	17	AA891856	Lactoferrin-derive
19	120	73.2	32	17	AA890610	Lactoferrin derive
20	120	73.2	32	17	AA887625	Lactoferrin-derive
21	120	73.2	32	18	AAW26154	Lactoferrin deriva
22	120	73.2	32	18	AAW14040	Anti-parasitic pep
23	120	73.2	32	19	AAW70314	Thrombus formation
24	120	73.2	32	19	AAW53229	Lactoferrin hydrol
25	120	73.2	32	21	AAW41294	Apoptosis inducer
26	120	73.2	32	21	AAW56543	Peptide SEQ ID NO:
27	120	73.2	38	14	AA84844	Lactoferrin-relate
28	120	73.2	38	15	AA848533	Lactoferrin derive
29	120	73.2	38	15	AA857465	Lactoferrin derive
30	120	73.2	38	16	AA884702	Bovine lactoferrin
31	120	73.2	38	16	AA880267	Anti-parasitic lac
32	120	73.2	38	17	AA891855	Lactoferrin-derive
33	120	73.2	38	17	AA890609	Lactoferrin derive
34	120	73.2	38	17	AA887624	Lactoferrin-derive
35	120	73.2	38	18	AAW26153	Lactoferrin deriva
36	120	73.2	38	18	AAW14039	Anti-parasitic pep
37	120	73.2	38	19	AAW70313	Thrombus formation
38	120	73.2	38	19	AAW53228	Lactoferrin hydrol
39	120	73.2	38	19	AAW41293	Apoptosis inducer
40	120	73.2	38	21	AAW56542	Peptide SEQ ID NO:
41	115	70.1	74	22	AAW51535	Recombinant lactof
42	105	64.0	51	17	AA891194	Lactoferrin decomp
43	105	64.0	52	16	AA887909	Bovine lactoferrin
44	100	61.0	36	14	AA84846	Lactoferrin-relate
45	100	61.0	36	17	AAW10517	Lactoferrin derive

ALIGNMENTS

RESULT 1	AA88341	AA88341 standard; peptide, 29 AA.
ID	AA88341	
XX	AA88341	
AC	AA88341	
XX	AA88341	
DT	14-JUL-2000 (first entry)	
DE	Peptide derived from an internal fragment of goat whey.	
XX		
XX	Peptide production; biological fluid; milk; whey; blood;	
KW	antibacterial peptide; lactoferrin; antiviral; antitumour activity.	
OS	Capra sp.	
XX		
FT	Key	Location/Qualifiers
FT	Disulfide-bond 6..23	/note="Optionally not present"
FT		
XX	WO200015655-A1.	
XX		
PD	23-MAR-2000.	
XX		
PF	15-SEP-1999;	99WO-EP07002.
XX		
PR	15-SEP-1998;	98EP-0203107.
XX		
PR	08-JUN-1999;	99EP-0201815.
XX		
PA	(NIZO-) NIZO FOOD RES.	
XX		
XX	Wiiser S, Reclio I;	
PI		
XX		

DR WPI; 2000-271377/23.

XX Novel process for producing peptides with e.g. antimicrobial activity
 PT from biological fluids such as milk, whey or blood comprises contacting
 PT fluid with chromatographic medium to adsorb peptide domain of interest
 PT

PS Claim 14; Page 22; 41pp; English.

CC This sequence represents a peptide derived from an internal fragment of
 CC goat whey. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.

XX Sequence 29 AA;

Query Match 100.0%; Score 164; DB 21; Length 29;

Best Local Similarity 100.0%; Pred. No. 8.4e-17;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PEMSKCYQWQRMKLGAPSTICIRRTSA 29
 |||
 DB 1 PEMSKCYQWQRMKLGAPSTICIRRTSA 29

RESULT 2

AAG80774 ID AAG80774 standard; Protein; 708 AA.

AC AAG80774;

DT 19-APR-2002 (first entry)

DE Goat lactoferrin-associated protein #1.

KM Primer; lactoferrin; antibiotic; goat.

OS Capra sp.

FX Key Location/Qualifiers

FT Misc-difference 360 /note= "Encoded by CGG"

PN KR98043944-A.

PD 05-SEP-1998.

PF 05-DEC-1996; 96KR-0061925.

PR 05-DEC-1996; 96KR-0061925.

PA (KOAD) KOREA ADV INST SCI & TECHNOLOGY.

PI YOO DY, Nam MS, Lee TH, Yoo SR, Kim SJ, Lee GG;

DR WPI; 1999-465631/39.

DR N-PSDB; ABA97187.

PT Antibiotic lactoferrin derived from Korean black goat and gene thereof -
 PT NoAbstract

PS Disclosure; Page 6-9; 11pp; Korean.

XX This sequence represents the goat antibiotic lactoferrin described in the
 CC disclosure of the invention.

XX Sequence 708 AA;

Query Match 100.0%; Score 164; DB 20; Length 708;

Best Local Similarity 100.0%; Pred. No. 2.6e-15;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PEMSKCYQWQRMKLGAPSTICIRRTSA 29
 |||
 DB 33 PEMSKCYQWQRMKLGAPSTICIRRTSA 61

RESULT 3

AAG80775 ID AAG80775 standard; Protein; 708 AA.

AC AAG80775;

DT 19-APR-2002 (first entry)

DE Goat lactoferrin-associated protein #2.

KM Primer; lactoferrin; antibiotic; goat.

OS Capra sp.

FX Key Location/Qualifiers

FT Misc-difference 360 /note= "Encoded by CGG"

PN KR98043944-A.

PD 05-SEP-1998.

PF 05-DEC-1996; 96KR-0061925.

PR 05-DEC-1996; 96KR-0061925.

PA (KOAD) KOREA ADV INST SCI & TECHNOLOGY.

PI YOO DY, Nam MS, Lee TH, Yoo SR, Kim SJ, Lee GG;

DR WPI; 1999-465631/39.

PT Antibiotic lactoferrin derived from Korean black goat and gene thereof -
 PT NoAbstract

PS Disclosure; Page 6-9; 11pp; Korean.

CC This sequence represents the goat antibiotic lactoferrin described in the
 CC disclosure of the invention.

XX Sequence 708 AA;

Query Match 99.4%; Score 163; DB 20; Length 708;

Best Local Similarity 96.6%; Pred. No. 3.6e-13;

Matches 28; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PEMSKCYQWQRMKLGAPSTICIRRTSA 29
 |||
 DB 33 PEMSKCYQWQRMKLGAPSTICIRRTSA 61

RESULT 4

AAV83112 ID AAV83112 standard; peptide; 25 AA.

AC AAV83112;

DT 24-JUL-2000 (first entry)
 XX Synthetic lactoferrin fragment.
 DE Lactoferrin; antimicrobial; cytostatic; infection; tumour;
 XX Sterilisation; wound healing; spermicide; goat.
 KM Synthetic.
 OS Capra hircus.
 XX WO200012541-A2.
 PN 09-MAR-2000.
 PD 31-AUG-1999; 99MO-GB02850.
 XX 28-AUG-1998; 98GB-0018938.
 PR (ALPH-) ALPHARMA AS.
 PA (GARD/) GARDNER R.
 XX Svendsen JS, Rekdal O, Sveinbjornsson B, Vorland L;
 FI WPI, 2000-270793/23.
 DR Cytotoxic peptides useful as medicament for treating tumors and
 XX bacterial infections, comprises one or more non-genetic bulky and
 PT lipophilic amino acids
 PS Example 1; Figure 1; 114pp; English.
 XX Cytotoxic 7-25 mer lactoferrin peptides which comprise three or more
 CC cationic residues and have one or more non-genetic bulky and
 CC lipophilic amino acids have cytostatic and antimicrobial activity.
 CC The peptides are useful as medicament for treating bacterial
 CC infections and tumours. They are also is useful for sterilising
 CC agents for materials susceptible to microbial contamination. They may
 CC also be used as promoters of wound healing and spermicides. The small
 CC size of the peptides gives them an increased half life and allows
 CC efficient biodelivery. The peptides can be administered without need
 CC for an injection, such as by inhalation or by absorption across the
 CC blood capillaries of the nasal passages. This synthetic lactoferrin
 CC peptide corresponds to amino acids 17-41 of caprine lactoferrin.
 XX SQ Sequence 25 AA;
 Query Match 82.9%; Score 136; DB 21; Length 25;
 Best Local Similarity 96.0%; Pred. No. 8.8e-13;
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 4 SKCYQWRMRKLGAPSTICIRRTS 28
 Db 1 SKCYQWRMRKLGAPSTICVRRTS 25
 RESULT 5
 ID AAY85013 standard; peptide; 25 AA.
 XX AAY85013;
 AC AAY85013;
 XX 20-JUN-2000 (first entry)
 DT Caprine lactoferrin peptide LFC(17-41) amino acid sequence.
 XX Lactoferrin; lactoferricin; antibacterial; bacterial infection;
 KW bacterial growth; solid tumour; treatment; goat.
 XX Capra sp.
 OS WO200012542-A2.
 XX 09-MAR-2000.
 PD

XX 31-AUG-1999; 99MO-GB02851.
 XX 28-AUG-1998; 98GB-0018938.
 PR (ALPH-) ALPHARMA AS.
 PA (GARD/) GARDNER R.
 XX Svendsen JS, Rekdal O, Sveinbjornsson B, Vorland L;
 FI WPI, 2000-256582/22.
 DR Modified cytotoxic lactoferrin peptide useful for preparing
 PT pharmaceutical composition against bacterial infections and
 PT for inhibiting bacterial growth
 XX Example 1; Fig 1; 112pp; English.
 PS This sequence represents a caprine lactoferricin peptide. Lactoferrin is
 CC a weak iron transporter which also shows weak antibacterial effect.
 CC Lactoferricin B (LFB) is a peptide produced when bovine lactoferrin is
 CC digested with pepsin. LFB is more active than lactoferrin. The two
 CC tryptophan residues at positions 6 and 8 in LFB 17-31 cannot be
 CC substituted with alanine without loss of antibacterial activity. The
 CC invention relates to a modified cytotoxic lactoferrin peptide of 7-25
 CC amino acids in length with three cationic residues and one or more extra
 CC bulky and lipophilic amino acids, than its native peptide. The invention
 CC also relates to the esters, amides, salts and cyclic derivatives of the
 CC modified peptide. The modified lactoferrin peptides of the invention are
 CC useful for treating bacterial infections and tumours, for preparing
 CC medicaments and for inhibiting bacterial growth. LFB and its fragments
 CC are useful for manufacturing medicaments for treating solid tumours. The
 CC modified peptides have a very small size, which is suitable for
 CC biodelivery. Due to the smaller size, the circulating half-life of the
 CC peptide is increased as they are less vulnerable to endopeptidases. The
 CC observed cytolytic effect of the modified lactoferrin peptides in tumours
 CC is not species specific and they have greater utility in treating human
 CC tumours.
 XX SQ Sequence 25 AA;
 Query Match 82.9%; Score 136; DB 21; Length 25;
 Best Local Similarity 96.0%; Pred. No. 8.8e-13;
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 4 SKCYQWRMRKLGAPSTICIRRTS 28
 Db 1 SKCYQWRMRKLGAPSTICVRRTS 25
 RESULT 6
 ID AAY49270 standard; protein; 689 AA.
 XX AAY49270;
 AC AAY49270;
 XX 07-FEB-2000 (first entry)
 DT Lactoferrin sequence.
 XX Adhesion inhibition; enteropathogenic; Escherichia coli; lactoferrin;
 KW medicament; drink; feed; food poisoning.
 XX Unidentified.
 OS JP11292789-A.
 XX 26-OCT-1999.
 PD 03-APR-1998; 98JP-0107167.
 XX 03-APR-1998; 98JP-0107167.
 PR 03-APR-1998; 98JP-0107167.
 XX

(MEIP) MEIJI MILK PROD CO LTD.

WIPI; 2000-018674/02.

Adhesion inhibitory compositions of enteropathogenic E. coli to cells - used in medicines and foodstuffs

Disclosure; Fig 4; 7pp; Japanese.

The invention provides an adhesion inhibitory composition of enteropathogenic Escherichia coli to cells. The composition contains a lactoferrin degraded matter. The composition is useful as a medicament, a drink and food or a feed. The composition is effective for prevention and/or treatment of infectious food poisoning caused by enteropathogenic E. coli.

Sequence 689 AA;

Query Match Best Local Similarity 74.4%; Score 122; DB 21; Length 689;
Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0

1 PEWSKYQWOFMRMKLGAPSITCIRRTSA 29
||| ||| : ||| : ||| ||| |||
14 PFWFKCRMQMGRMKKLGAPSITCVRRAPA 42

RESULT 7
ID AM09343 standard; Protein; 708 AA.
XX AC AAM09343;
XX DT 25-MAR-2003 (updated)
XX DT 18-MAR-1997 (first entry)

Bovine lactoferrin.

Human lactoferrin; iron-binding glycoprotein; milk; secretion; fungus; transferrin; bactericidal activity; prostate; expression system; primer; PCR; polymerase chain reaction; amplification; signal peptide; antiviral; alpha-amylase; Aspergillus oryzae; nutrition; bovine.

Bos taurus .

US5571691-A.
XX PD 05-NOV-1996.
XX PP 28-OCT-1993; 93US-0145681.
XX PR 28-OCT-1993; 93US-0145681.
XX PR 05-MAY-1989; 89US-0348270.
XX PR 28-SEP-1989; 89US-0413880.
XX PR 24-APR-1992; 92US-0873304.
XX PR 27-OCT-1992; 92US-0967947.

(BAYU) BAYLOR COLLEGE MEDICINE.
XX PA Connely OM, Headon DR, May GS, O'Malley BW,
XX PI WIPI; 1996-505400/50.
XX DR N-PDB: MAT46031.

New DNA encoding human lactoferrin, its natural alleles and substitution analogues - useful e.g. for preventing iron deficiency and as antiviral/antimicrobial agent

Disclosure; Column 31-34; 92pp; English.

This is the amino acid sequence of bovine lactoferrin, an iron-binding glycoprotein found in milk and other secretions and body fluids. Fragments of the protein are also known to have biological activity e.g.

```

CC the N-terminal portion of the protein has a bactericidal activity. The
CC gene was used to construct a fusion protein in which the native
CC lactoferrin signal peptide was replaced by the alpha-amylase II signal
CC peptide. The novel construct was then expressed in Apbergillus oryzae.
CC The protein can be used for antibacterial and antiviral activities as
CC well as an iron-carrying protein for nutritional or therapeutic
CC applications.
CC (Updated on 25-MAR-2003 to correct PF field.)
CC
XX
SQ Sequence 708 AA;
XX
Query Match 74.4%; Score 122; DB 17; Length 708;
Best Local Similarity 72.4%; Pred. No. 3.5e-09;
Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 PEMSECYQMRRMRKAGPSITCIRNTSA 29
   |||||:|||||:|||||:|||||:
Db 33 PEMFKCRMQMRMKKLAGPSITCYVRAFA 61
XX
RESULT 8
XX AAM57318
XX ID AAM57318 standard; Protein; 708 AA.
XX AAM57318;
XX AA57318;
XX DT 10-AUG-1998 (first entry)
XX DE Bovine lactoferrin.
XX
XX Bovine; lactoferrin; recombinant; therapeutic; nutritional; iron;
XX Fe binding site; bacteria; bactericidal; milk.
XX
XX Bos taurus.
XX
XX US5766939-A.
XX
XX PD 16-JUN-1998.
XX
XX PF 30-MAY-1995; 95US-0453703.
XX
XX PR 28-OCT-1993; 93US-0145681.
XX PR 05-MAY-1989; 89US-0348270.
XX PR 24-APR-1992; 92US-0873304.
XX PR 27-OCT-1992; 92US-0967947.
XX PR 30-MAY-1995; 95US-0453703.
XX
XX (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
XX Comeely OM, Heaton DR, May GS, O'Malley BM;
XX WPI, 1998-361705/31.
XX DR N-PSDB; AAV30771.
XX
XX Plasmids containing human lactoferrin DNA - for recombinant
XX production of the enzyme, especially fragments having bactericidal
XX activity
XX
XX Example 13; Fig 14; 92pp; English.
XX
XX The present sequence represents bovine lactoferrin. A plasmid has been
XX developed which is suitable for the expression of a human lactoferrin,
XX or an iron-binding lobe of lactoferrin, in a transformed prokaryotic
XX host cell. The plasmid comprises a DNA sequence encoding a naturally
XX occurring human lactoferrin protein or an iron-binding lobe of
XX lactoferrin and further comprises transcriptional and translational
XX regulatory elements capable of regulating the expression of the
XX lactoferrin-encoding DNA sequence in the transformed host cell. The
XX plasmid is useful for producing recombinant human lactoferrin proteins
XX in bacteria. Lactoferrin is a 78 kDa iron-binding glycoprotein found
XX in milk and other secretory fluids. It is involved in iron transfer
XX and delivery in mammals. It has been implicated as a resistance
XX factor in suckled new born infants against enteritis infections; the
XX

```


XX	Lactoferrin; antibiotic; chelate; mastitis; bowel disorder;
KW	disease; bacteria; yeast; fungi; disinfection; drug; foodstuff;
KX	cosmetic; toiletries.
OS	Unidentified.
XX	
Key	Location/Qualifiers
PH	Disulfide bond 10..27
PX	
PN	WO9314640-A1.
PD	
PP	05-AUG-1993.
XX	
PE	30-NOV-1992; 92WO-JP01563.
XX	
PR	23-JAN-1992; 92JP-0032660.
PR	11-MAR-1992; 92JP-0052943.
PR	30-SEP-1992; 92JP-0262143.
PR	30-SEP-1992; 92JP-0262359.
PA	(MORG) MORINAGA MILK IND CO LTD.
XX	
PI	Belamy W, Fukuwatari Y, Kawase K, Shimamura S;
PI	Takase M, Tokitay, Tomita M, Wakabayashi H, Yamauchi K;
XX	WPI; 1993-258265/32.
DR	
XX	
PT	Antibacterial agent comprising decomposition products of
PT	lactoferrin - with chelate e.g. EXTRA alcohol and/or antibiotic
PT	e.g. penicillin, also useful against yeast and fungi
XX	
PS	Disclosure; Page 87; 100pp; Japanese.
XX	
CC	Lactoferrin-related peptides are used in new antibacterial compn.
CC	The compn. is highly effective against a broad range of bacteri,
CC	yeasts and fungi. It can be used therapeutically (internal and
CC	external application), e.g. for mastitis, bowel disorders, urinary
CC	infections, etc. It can also be used for the disinfection and
CC	protection of drugs, foodstuffs, cosmetics and toiletries and
CC	household items (such as kitchen towels and toilet paper).
CC	(Updated on 09-JAN-2003 to add missing OS field.)
CC	(Updated on 25-MAR-2003 to correct PN field.)
XX	
SQ	Sequence 32 AA:
Query Match	73.2%; Score 120; DB 14; Length 32;
Best Local Similarity	76.9%; Pred. No. 2.5e-10;
Matches 20; Conservative 3; Mismatches 3; Indels 0; Gaps 0;	
OY	1 PEWSKYCWOMRRTKGAPSTICRR 26
DB	5 PEFKCRKRWOMRKTKGAPSTICVR 30
RESULT 14	
ID AAR48534	AAR48534
XX AAR48534;	
AC	
DT	25-MAR-2003 (updated)
DT	10-AUG-1994 (first entry)
XX	
DE	Lactoferrin derived peptide #28.
XX	
KM	Decomposition; lactoferrin; digestion; enzyme; trypsin;
KM	antioxidant; oxidation; inhibitor; vitamin B; ascorbic acid;
KM	vitamin A; beta-carotene; superoxide dismutase; coenzyme Q;
KM	lipid oxidation; foodstuff; drugs; health food; toiletries; cosmetics.
XX	
OS	Bos taurus.
XX	

XX	MOM9403555-A1.
PD	17-FEB-1994.
XX	
PF	04-AUG-1993; 93WO-JP01090.
XX	
PR	07-AUG-1992; 92JP-0211335.
XX	
PA	(MORG) MORINAGA MILK IND CO LTD.
PJ	Bellmy WR, Fukuwatari Y, Kawase K, Shimamura S;
PI	Takase M, Tokiday, Tomita M, Wakabayashi H, Yamauchi K;
DR	WPI, 1994-065650/08.
XX	
PT	Antioxidant peptide lactoferrin decomposition product - prevents oxidation of lipid(s) in foodstuffs and drugs without affecting their taste
XX	
PS	Claim 3; Page 39; 47p; Japanese.
XX	
CC	The sequences given in AR48507-37 are peptides derived by the decomposition of lactoferrin, pref. by digestion with an enzyme, eg. pepsin or trypsin. These peptides may be used in an antioxidant composition which may also contain an oxidation inhibitor such as vitamin B, ascorbic acid, vitamin A, beta-carotene, superoxide dismutase or coenzyme Q. The antioxidant prevents lipid oxidation in foodstuffs, drugs, health foods, toiletries and cosmetics. (updated on 25-MAR-2003 to correct PN field.)
CC	
CC	
Sequence	32 AA;

```

Query Match Similarity 73.2%; Score 120; DB 15; Length 32;
Best Local Similarity 76.9%; Pred. No. 2,5e-10;
Matches 20; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY      1 PEMSRCTQWQRRMRKLGAPSTTCRR 26
      ||| ||| : ||| ||| ||| ||| |||
Db      5 PEMFKCRRWQWRMKKLGAPSTTCVR 30

RESULT 15
AAR57466
ID AAR57466 strand: Protein; 32 AA.
AC AAR57466;
XX
XX 28-FEB-1995 (first entry)
XX
XX
XX Lactoferrin derived peptide #29.
XX
XX Lactoferrin; Chemical; enzymatic; hydrolysis; antimicrobial;
XX
XX antiseptic; ischaemic disease.
XX
XX Mus musculus.
XX
XX JP06172200-A.
XX
XX
XX 21-JUN-1994.
XX
XX
XX 08-DEC-1992; 92JP-0327738.
XX
XX
XX 08-DEC-1992; 92JP-0327738.
XX
XX
XX (MORG ) MORINAGA MILK IND CO LTD.
XX
XX WPI; 1994-238662/29.
XX
XX
XX Brain protectant for preventing ischaemic diseases without side
XX
XX effects - comprising 31 specified peptide(s), prepd. by
XX
XX lactoferrin hydrolysis
XX
XX
XX Disclosure; Page 10; 11pp; Japanese.

```

XX The sequences given in AARS7438-68 represent fragments of lactoferrin
CC which were derived from the full length protein by chemical or enzyme
CC hydrolysis. These peptides have brain protecting properties, as
CC well as anti-microbial activity. Compositions containing these
CC peptides may be prepared with out the addition of antiseptics, and
CC may be administered at doses of at least 10 mg for parenteral
CC administration and 100 mg for oral administration. These peptides
CC are stable, heat resistant, water soluble and may be used for the
CC prevention of ischemic diseases without side effects.

XX Sequence 32 AA;

Query Match 73.2%; Score 120; DB 15; Length 32;
Best Local Similarity 76.9%; Pred. No. 2.5e-10;
Matches 20; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 PEMSKCYQWRMRKLGAPSTICRR 26
||| ||| : ||| ||| ||| : |||
Db 5 PEMFKCRRWQWRMRKLGAPSTICRR 30

Search completed: July 30, 2003, 16:23:28
Job time : 57.1369 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 27.1429 Seconds

(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-7

Sequence: 1 TORXNGFRVPLARE 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: SPTRMBL_23:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*
16: sp_virus:*
17: sp_bacteriap:*
17: sp_archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	712	6 Q9MZY2	Q9mzy2 ovis aries
2	55	67.1	711	6 Q9GTM6	Q9gjm6 camelus dromedarius
3	45	54.9	710	11 Q8R481	Q8r481 mesocricetus
4	44	53.7	359	16 Q9RMV1	Q9rmv1 vibrio cholerae
5	44	53.7	667	3 Q8X0K1	Q8x0k1 neurospora crassa
6	43	52.4	725	13 Q8QFX3	Q8qfx3 xenopus laevis
7	43	52.4	1134	16 Q9RUU0	Q9ruu0 streptomyces
8	42	51.2	710	11 Q9IWA0	Q9iwa0 mus musculus
9	42	51.2	1394	16 Q8G3M2	Q8g3m2 bifidobacterium
10	41	50.0	138	16 Q8PFA8	Q8pfa8 xylella fastidiosa
11	41	50.0	414	5 Q21764	Q21764 caenorhabditis elegans
12	41	50.0	417	16 Q8FLU8	Q8flu8 corynebacterium
13	41	50.0	631	16 Q912G3	Q912g3 pseudomonas aeruginosa
14	41	50.0	876	10 Q64483	Q64483 arabidopsis thaliana
15	40	48.8	1333	16 Q8NT25	Q8nt25 corynebacterium
16	40	48.8	1333	16 Q8PS96	Q8ps96 corynebacterium

17	39.5	48.2	1171	16 Q92PS9	Q92ps9 rhizobium meliloti
18	39	47.6	76	4 Q9Y3L1	Q9y3l1 homo sapiens
19	39	47.6	95	15 Q9QNT7	Q9qnt7 human immunodeficiency virus
20	39	47.6	126	2 Q8GUK8	Q8guk8 escherichia coli
21	39	47.6	128	17 Q59561	Q59561 pyrococcus furiosus
22	39	47.6	182	5 Q9N439	Q9n439 caenorhabditis elegans
23	39	47.6	226	16 Q8KF28	Q8kf28 chlorobium thiosulfatophilum
24	39	47.6	235	10 Q9C653	Q9c653 arabidopsis thaliana
25	39	47.6	314	10 Q8VZM8	Q8vzm8 arabidopsis thaliana
26	39	47.6	320	5 Q9SU84	Q9su84 diatomosphaera
27	39	47.6	332	17 Q97917	Q97917 thermoplasma
28	39	47.6	376	16 Q8XNR9	Q8xnr9 rickettsia
29	39	47.6	418	16 Q8EL08	Q8el08 oceanobacillus
30	39	47.6	625	4 Q8N4A6	Q8n4a6 homo sapiens
31	39	47.6	629	16 Q8Z050	Q8z050 anabaena
32	39	47.6	656	4 Q8WV21	Q8wv21 homo sapiens
33	39	47.6	684	10 Q9ZV55	Q9zv55 arabidopsis thaliana
34	39	47.6	1029	4 Q8TBX2	Q8tbx2 homo sapiens
35	39	47.6	1156	4 Q8IUL8	Q8iul8 homo sapiens
36	39	47.6	1281	12 Q854Z5	Q854z5 rat cytochrome
37	39	47.6	1727	4 Q9UPQ9	Q9upq9 homo sapiens
38	38.5	47.0	310	17 Q8UIF1	Q8uif1 pyrococcus
39	38	46.3	123	16 Q8YI18	Q8yi18 anabaena
40	38	46.3	172	16 P96883	P96883 mycobacterium
41	38	46.3	185	16 Q9X883	Q9x883 streptomyces
42	38	46.3	204	4 Q9NTP6	Q9ntp6 homo sapiens
43	38	46.3	251	2 Q9AMH3	Q9amh3 streptomyces
44	38	46.3	251	2 Q9F5U0	Q9f5u0 streptomyces
45	38	46.3	268	17 Q8ITV3	Q8itv3 methanobrevibacter

ALIGNMENTS

RESULT 1

ID	Q9MZY2	PRELIMINARY;	PRT;	712 AA.
AC	Q9MZY2			
DT	01-OCT-2000 (TREMURel. 15, Created)			
DT	01-OCT-2000 (TREMURel. 15, Last sequence update)			
DT	01-JUN-2002 (TREMURel. 21, Last annotation update)			
DE	Alway lactoperoxidase.			
OS	Ovis aries (sheep).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Caprinae; Ovis.			
OX	NCBI_TaxID=9940;			
RN	[1]			
RP	SEQUENCE FROM N. A.			
RC	TISUB=Tracheal mucosa;			
RK	MEDLINE=20236390; PubMed=10837362;			
RA	Gerson C., Sabater J., Scurt W., Torbatti A., Coffey R., Abraham J.W.,			
RA	lauredo I., Forteza R., Wanner A., Salathe M., Abraham W., Conner G.,			
RT	"The Lactoperoxidase System Functions in Bacterial Clearance of			
RT	Always."			
RT	Am. J. Respir. Cell Mol. Biol. 22:665-671 (2000).			
DR	EMBL; AF027970; AAF74979.1; -			
DR	HSSP; P05164; 1CXP			
DR	InterPro; IPR002007; Anni_peroxidase.			
DR	Pfam; PF03096; An_peroxidase; 1.			
DR	PRINTS; PR00457; ANPEROXIDASE.			
DR	SEQUENCE 712 AA; 80486 MW; 5D3455943D6426F2 CRC64;			

Query Match	100.0%	Score 82;	DB 6;	Length 712;
Best local similarity	100.0%	Pred. No. 4.5e-06;		
Matches 16;	Conservative 0;	Mismatches 0;	Gaps 0;	
Indels 0;				
QY	1 TORXNGFRVPLARE 16			
DB	179 TORXNGFRVPLARE 194			

Query Match 53.7%; Score 44; DB 3; Length 667;
 Best Local Similarity 69.2%; Pred. No. 28;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 TORKTRNGFRVPL 13
 ||| |||||
 DB 284 TSKRLGGGFRVPL 296

RESULT 6

OQ8QFX3 PRELIMINARY; PRT; 725 AA.
 AC OQ8QFX3;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Peroxidase 2'.
 GN POX2'.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodidae; Xenopus.
 NC NCB1_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smith S.J., Kotecha S., Towers N., Mohun T.J.;
 RT "Xlure-1 and XPOX2 peroxidase expression reveal the site of embryonic
 RT myeloperoxidase in Xenopus."
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY069942; AL55400.1; -.
 DR InterPro; IPR002007; Anim_peroxidase.
 DR Pfam; PF03098; An_peroxidase; 1.
 SQ SEQUENCE 725 AA; 82417 MW; 8546878368C4947F CRC64;

Query Match 52.4%; Score 43; DB 13; Length 725;
 Best Local Similarity 53.3%; Pred. No. 46;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 TORKTRNGFRVPLAR 15
 ||| |||||
 DB 187 TENRRINGFPLPLAR 201

RESULT 7

OQ9RJUO PRELIMINARY; PRT; 1134 AA.
 AC OQ9RJUO;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Possible DNA-binding protein.
 GN SC00370 OR SCF41.29C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycinae; Streptomycetaceae; Streptomyces.
 NC NCB1_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-A3(2);
 RC Saunders D.C., Harris D.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RL Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
 RN Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RP [3]
 RC STRAIN-A3(2);
 RC STRAIN-A3(2);
 RC MEDLINE=97000351; Pubmed=8843436;
 RA Redenbach M., Kleiser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RA "A set of ordered cosmids and a detailed genetic and physical map for

the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RT Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RX MEDLINE=21996410; Pubmed=12000953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.B., Quail M.A., Kleiser H.,
 RA Harper D., Bauman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Frazer A., Goble A., Hidalgo J., Hornby T., Howarth S.,
 RA Huang C.-H., Kleiser J., Larke B., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietorek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)";
 RL Nature 417:141-147(2002).
 DR EMBL; AL393105; CAB55732.1; -.
 DR InterPro; IPR006311; Tat.
 DR TIGRFAMs; TIGR01409; Tat_signal_seq; 1.
 DR DNA-binding; Complete proteome.
 SQ SEQUENCE 1134 AA; 121689 MW; 89B689CD455C4520 CRC64;

Query Match 52.4%; Score 43; DB 16; Length 1134;
 Best Local Similarity 64.3%; Pred. No. 72;
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 RKTNRNGFRVPLARE 16
 ||| |||||
 DB 913 RATNGVGRVPLDLE 926

RESULT 8

OQ1WAO PRELIMINARY; PRT; 710 AA.
 AC OQ1WAO;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical 79.9 kDa protein.
 GN LPO.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAUSBERG R.;
 RC Tissue=Salivary gland;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC016212; AAH16212.1; -.
 DR MGD; MGI:1923363; Lpo.
 DR InterPro; IPR002007; Anim_peroxidase.
 DR Pfam; PF03098; An_peroxidase; 1.
 DR PRINTS; PR00457; ANPEROXIDASE.
 KW Hypothetical protein
 SQ SEQUENCE 710 AA; 79922 MW; EB9BCF12472EC5DA CRC64;

Query Match 51.2%; Score 42; DB 11; Length 710;
 Best Local Similarity 56.2%; Pred. No. 67;
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 TORKTRNGFRVPLARE 16
 ||| |||||
 DB 177 TPKMGNGFRVPLPRE 192

RESULT 9

OQ83M2 PRELIMINARY; PRT; 1394 AA.
 ID OQ83M2;
 AC OQ83M2;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMblrel. 23, last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, last annotation update)
 DE Possible helicase.
 GN BL1736.
 OS Bifidobacterium longum.
 OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
 OC Bifidobacteriaceae; Bifidobacterium.
 NCBI_TaxID=216816;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCC 2705;
 RX MEDLINE=22294977; PubMed=12381787;
 RA Scheil M.A., Karamitancou M., Snel B., Villanova D., Berger B.,
 RA Pessi G., Zwaalen M.-C., Desiere F., Bork P., Delley M.,
 RA Pridmore R.D., Arigoni F.;
 RT "The genome sequence of Bifidobacterium longum reflects its adaptation
 RT to the human gastrointestinal tract."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
 DR EMBL; AB014807; AANZ520.1; -.
 KW Helicase; Complete proteome.
 SQ SEQUENCE 1394 AA; 151986 MW; 7393P1750312D9P3 CRC64;

Query Match 51.2%; Score 42; DB 16; Length 1394;
 Best Local Similarity 61.5%; Pred. No. 1.4e+02;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 KTRNGFRVPLARE 16
 Db 694 QTVNGFRTPARK 706

RESULT 10
 O9PF48 PRELIMINARY; PRT; 138 AA.
 ID O9PF48.
 AC O9PF48.
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, last annotation update)
 DE Hypothetical protein XF0830.
 GN XF0830.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9a5c;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach P.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
 RA Colauto N.B., Colombo C., Costa F.P., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Fulton L.R.,
 RA Gantier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kirijsima J.P.,
 RA Krieger J.B., Kiriama E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
 RA Marquis M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto R.R., Pereira G.A.G., Pereira H.A., Jr., Pequeiro J.B.,
 RA Queiroz R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawaeki H.B.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.P., Truffi D., Teal S.M., Tsuchiko M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,

RA Zago M.A., Zatz M., Weidman J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa."
 RL Nature 406:151-159(2000).
 DR EMBL; AE003922; AAF83640.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 138 AA; 15117 MW; 1A9397C765730FC1 CRC64;

Query Match 50.0%; Score 41; DB 16; Length 138;
 Best Local Similarity 64.3%; Pred. No. 19;
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 QKTRNGFRVPLAR 15
 Db 123 QKTRNGFRVPLAR 136

RESULT 11
 ID Q21764 PRELIMINARY; PRT; 414 AA.
 AC Q21764;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, last annotation update)
 DE R05H5.4 protein.
 GN R05H5.4
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloditerinae; Caenorhabditis.
 NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA McMurray A.A.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP MEDLINE=99069613; PubMed=9851916;
 RX none;
 RA "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL; Z48795; CAA88727.1; -.
 DR HSSP; P12931; 1FMK.
 DR WormPep; R05H5.4; CE02291.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR000980; SH2.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00017; SH2; 1.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00252; SH2; 1.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 414 AA; 47188 MW; E51F5CD772E9A908 CRC64;

Query Match 50.0%; Score 41; DB 5; Length 414;
 Best Local Similarity 63.6%; Pred. No. 58;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 QKTRNGFRVPLAR 12
 Db 327 KKKTRAGVRLP 337

RESULT 12
 ID Q8FLU8 PRELIMINARY; PRT; 417 AA.
 AC Q8FLU8.
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, last sequence update)


```

DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)
DE Putative L-lactate dehydrogenase.
GN CE2762.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OK NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11199;
RA Kawarayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima U., Itoh T., Yamagishi A., Nishio Y.,
RA Ueda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005223; BAC19572.1;
KW Complete proteome.
SQ SEQUENCE 417 AA; 45519 MW; 127DD710E3C7A500 CRC64;

QY Query Match 50.0%; Score 41; DB 16; Length 417;
Best Local Similarity 70.0%; Pred. No. 59;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 3 RKTNGFRVPL 12
197 RDTNGFSIP 206

RESULT 13
ID Q912G3 PRELIMINARY; PRT; 631 AA.
AC Q912G3;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DB 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein PA1941.
GN PA1941.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OK NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
MEDLINE=2043733; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lathig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Siler M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004620; AAG05329.1;
DR InterPro; IPR000345; CytC_heme_bind.
DR PROSITE; PS00190; CYTOCHROME_C_2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 631 AA; 70904 MW; E72A4692AEB75F1 CRC64;

QY Query Match 50.0%; Score 41; DB 16; Length 611;
Best Local Similarity 58.3%; Pred. No. 90;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 2 ORKTRNGFRVPL 13
474 OQKTRNGFDLPI 485

RESULT 14
ID O64483 PRELIMINARY; PRT; 876 AA.
AC O64483;

```

```

DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Putative receptor-like protein kinase (Senescence-induced receptor-
DE like serine/threonine kinase).
GN At2G19190 OR SIKK.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OK NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldhym T.V.,
RA Bueli C.R., Ketchum K.A., Lee J.U., Honing C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.B., Mayhew D., Talton L.J., Gail J.B.,
RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,
RA Salzberg S.L., Frazer C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Robatzek S., Somssich I.E.;
RT "Targets of AtWRKY regulation during plant senescence and pathogen
RT defense.";
RL Genes Dev. 0:0-0(2002).
CC -1 SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AC002392; AAD12037.1;
DR EMBL; AF486619; AAL92103.1;
DR InterPro; IPR001611; LRR.
DR InterPro; IPR007090; LRR_plant.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00560; LRR_3.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PSS0502; LRR_PS; 1.
DR PROSITE; PSS0107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PSS0108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Receptor; Serine/threonine-protein kinase;
KW Transferrase.
SQ SEQUENCE 876 AA; 98711 MW; EEB1C13E335BC786 CRC64;

QY Query Match 50.0%; Score 41; DB 10; Length 876;
Best Local Similarity 46.2%; Pred. No. 1,3e+02;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

DB 1 TQKTRNGFRVPL 13
510 TKKKNKNGYIPL 522

RESULT 15
ID Q8NT25 PRELIMINARY; PRT; 1333 AA.
AC Q8NT25;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE DNA-directed RNA polymerase beta' subunit/160 kDa subunit (split gene
DE in archaea and eukaryotes) (EC 2.7.7.6).
GN CGL0489.

```

OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
 OK NCBI_taxid=1718;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032."
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP005275; BAB97882.1; -;
 DR InterPro; IPR007022; RNA_pol_A.
 DR InterPro; IPR007080; RNA_pol_Rpdl_1.
 DR InterPro; IPR007066; RNA_pol_Rpdl_3.
 DR InterPro; IPR007083; RNA_pol_Rpdl_4.
 DR InterPro; IPR007081; RNA_pol_Rpdl_5.
 DR Pfam; PF04997; RNA_pol_Rpdl_1; 1.
 DR Pfam; PF00623; RNA_pol_Rpdl_2; 1.
 DR Pfam; PF04983; RNA_pol_Rpdl_3; 1.
 DR Pfam; PF05000; RNA_pol_Rpdl_4; 1.
 DR Pfam; PF04998; RNA_pol_Rpdl_5; 1.
 KW DNA-directed RNA polymerase; Transferase; Nucleotidyltransferase;
 KW Complete proteome.
 SQ SEQUENCE 1333 AA; 147294 MW; 0F6B89ADC719CF9 CRC64;

Query Match 48.8%; Score 40; DB 16; Length 133;
 Best Local Similarity 66.7%; Pred. No. 3e+02;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 5 TRNGFRVPLAE 16
 |||:|
 Db 903 TRCGVRVPVAAE 914

Search completed: July 30, 2003, 16:29:27
 Job time : 29.1429 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: July 30, 2003, 16:10:10 / Search time 5.2381 Seconds

(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-7

Sequence: 1 TORRNGFRVPLARE 16

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_41.1

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	712	1	PERL_BOVIN
2	59	72.0	712	1	PERL_HUMAN
3	51	62.2	718	1	PERL_MOUSE
4	46	56.1	716	1	PERL_MOUSE
5	44	53.7	421	1	ARGB_XYLFA
6	42	51.2	619	1	DNAX_MYCLE
7	42	51.2	715	1	PERE_HUMAN
8	41	50.0	488	1	HPIT_YEAST
9	41	50.0	876	1	SIRK_ARATH
10	39	47.6	320	1	BXD_DROME
11	38	46.5	463	1	VEIS_VZVD
12	38	46.3	491	1	ILVC_HAEIN
13	38	46.3	491	1	ILVC_PASMU
14	38	46.3	897	1	SVI_FERPE
15	38	46.3	1066	1	HGPG_HAEIN
16	38	46.3	1067	1	HGPG_HAEIN
17	38	46.3	1084	1	HGPG_HAEIN
18	38	46.3	4594	1	DYHC_HUMAN
19	38	46.3	4644	1	DYHC_HUMAN
20	38	46.3	4644	1	DYHC_MOUSE
21	37	45.1	242	1	Y503_VIBVU
22	37	45.1	252	1	HIS6_RHOSH
23	37	45.1	363	1	AG22_HUMAN
24	37	45.1	363	1	AG22_HUMAN
25	37	45.1	363	1	AG22_MOUSE
26	37	45.1	363	1	AG22_MOUSE
27	37	45.1	465	1	INNE_CARBL
28	37	45.1	642	1	NOGI_SCHPO
29	37	45.1	646	1	UVAB_METTH
30	37	45.1	745	1	PERM_HUMAN
31	37	45.1	1014	1	ACAA_ARATH
32	37	45.1	3027	1	POLG_PYFV1
33	36	43.9	44	1	RL34_BRUME

34	36	43.9	44	1	RL34_RALSO	O93389 ralsoria g
35	36	43.9	44	1	RL34_RHILLO	O96d90 rhizobium 1
36	36	43.9	242	1	Y538_VIBPA	O87886 vibrio para
37	36	43.9	251	1	HIS6_LISIN	O92888 listeria in
38	36	43.9	251	1	HIS6_LISMO	O89957 listeria mo
39	36	43.9	255	1	HIS6_NEIMA	O91v45 neisseria m
40	36	43.9	255	1	HIS6_NEIMA	O9K044 neisseria m
41	36	43.9	257	1	UNG_DICDI	P53766 dictyostel
42	36	43.9	271	1	HIS6_ARCRU	O29439 atcheaglob
43	36	43.9	446	1	TRME_CAUCR	O92d19 caulobacter
44	36	43.9	471	1	YIR1_DROME	P16424 drosophila
45	36	43.9	556	1	SRS4_CANAL	O42816 candida alb

ALIGNMENTS

RESULT 1
PERL_BOVIN STANDARD; PRT; 712 AA.
AC P80025;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactoperoxidase precursor (EC 1.11.1.7) (LPO).
GN LPO.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RA MEDLINE=91025552; PubMed=2222811;
RA Dull T.J., Oyeda C., Strosberg A.D., Nedwin G., Seilhamer J.J.;
RT "Molecular cloning of cDNAs encoding bovine and human
lactoperoxidase.";
RL Lactoperoxidase.";
RN [2]
RP DNA Cell Biol. 9:499-509(1990).
RP [2]
RP SEQUENCE OF 101-712.
RC TISSUE=Milk;
RX MEDLINE=91266958; PubMed=2050150;
RA Cals M.-M., Maillart P., Brignon G., Anglade P., Ribadeau-Dumas B.;
RT "Primary structure of bovine lactoperoxidase, a fourth member of a
mammalian heme peroxidase family.";
RL Eur. J. Biochem. 198:733-739(1991).
RN [3]
RP CIRCULAR DICHOISM ANALYSIS, AND PARTIAL SEQUENCE.
RX MEDLINE=20384171; PubMed=10924350;
RA Wacande S., Murata S., Kumura H., Nakamura S., Bollen A.,
RA Mogulilevsky N., Shimazaki K.-I.;
RT "Bovine lactoperoxidase and its recombinant: comparison of structure
and some biochemical properties.";
RL Biochem. Biophys. Res. Commun. 274:756-761(2000).
RN [4]
RP COVALENT HEME ATTACHMENT.
RX MEDLINE=97238804; PubMed=9083001;
RA Defillis G.D., Ozaki S.-I., Kuo J.M., Maltby D.A.,
RA Ortiz de Montellano P.R.;
RT "Allosteric processing of heme by lactoperoxidase produces the
native protein-bound prosthetic group.";
RL J. Biol. Chem. 272:8857-8860(1997).
RN [5]
RP COVALENT HEME ATTACHMENT.
RX TISSUE=Milk;
RC MEDLINE=98447642; PubMed=9774411;
RA Rae T.D., Golf H.M.;
RT "The heme prosthetic group of lactoperoxidase. Structural
characteristics of heme 1 and heme 1-peptides.";
RL J. Biol. Chem. 273:27968-27977(1998).
RN [6]
RP COVALENT HEME ATTACHMENT, AND MUTAGENESIS OF ASP-225 AND GLU-375.
RX MEDLINE=21481384; PubMed=11597409;

RA Suriano G., Matanabe S., Ghislaudi E.M., Bollen A., Ferrari R.P.,
 RA Moguilevsky N.,
 RT "Glut375Gln and Asp225Val mutants: about the nature of the covalent
 RT linkages between heme group and apo-protein in bovine
 RT lactoperoxidase.";
 RL Bioorg. Med. Chem. Lett. 11:2827-2831(2001).
 RN [7]
 RP COVALENT HEME ATTACHMENT, AND MUTAGENESIS OF ASP-225 AND GLU-375.
 RX MEDLINE=21850707; PubMed=11756449;
 RA Colas C., Kuo J.M., Ortiz de Montellano P.R.;
 RT "Asp-225 and Glu-375 in autocatalytic attachment of the prosthetic
 RT heme group of lactoperoxidase.";
 RL J. Biol. Chem. 277:7191-7200(2002).
 CC -1- FUNCTION: LPO IS AN ANTIMICROBIAL AGENT. IT IS THOUGHT TO HELP
 CC PROTECT THE UDDER FROM INFECTION AND PROMOTE GROWTH IN NEWMORN
 CC CALVES.
 CC -1- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
 CC -1- COFACTOR: Binds 1 prothème IX and 1 iron(III) ion.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. XPO SUBFAMILY.
 CC -1- DATABASE: NMBE=worthington-biochem manual;
 CC WWW="http://www.worthington-biochem.com/manual/L/LPO.html".
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M58150; AAA62714.1; -;
 DR PIR: A35828; A35828.
 DR HSRP: P05164; ICRP.
 DR InterPro: IPR002007; Anim_peroxidase.
 DR InterPro: IPR002016; Peroxidase.
 DR Pfam: PF03096; An_peroxidase; 1.
 DR PRINTS: PR00457; ANPEROXIDASE.
 DR PROSITE: PS00435; PEROXIDASE_1; FALSE NEG.
 DR PROSITE: PS00436; PEROXIDASE_2; FALSE NEG.
 DR PROSITE: PS50292; PEROXIDASE_3; 1.
 DR Oxidoreductase; Peroxidase; Iron; Heme; Glycoprotein; Milk; Signal;
 KW Antidiabetic.
 KM
 FT SIGNAL 1 22 POTENTIAL.
 FT PROPEP 23 100
 FT CHAIN 101 712 LACTOPEROXIDASE.
 FT ACT_SITE 226 226 DISTAL HISTIDINE (POTENTIAL).
 FT ACT_SITE 372 372 DISTAL ARGININE (POTENTIAL).
 FT BINDING 225 225 HEME (COVALENT).
 FT BINDING 375 375 HEME (COVALENT).
 FT METIL 468 468 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 394 394 A -> S.
 FT CONFLICT 449 449 N -> R (IN REF. 2).
 SQ SEQUENCE 712 AA; 80642 MW; 28BED4C0C82056D CRC64;
 Query Match 100.0%; Score 82; DB 1; Length 712;
 Best Local Similarity 100.0%; Pred. No. 8.5e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TORKTRNGFRVFLARE 16
 DB 179 TORKTRNGFRVFLARE 194

AC P22079; O13408;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DB Lactoperoxidase precursor (BC 1.11.1.7) (LPO) (Salivary peroxidase)
 DE (SPC).
 GN LPO.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_Taxid=9606;
 OK [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE-Submandibular gland;
 RX MEDLINE=97082979; PubMed=8964511;
 RA Kiser C., Caterina J., Engler J.A., Rahemtulla B., Rahemtulla F.;
 RT "Cloning and sequence analysis of the human salivary peroxidase-
 RT encoding cDNA.";
 RL Gene 173:261-264(1996).
 RN [2]
 RP SEQUENCE OF 389-712 FROM N.A.
 RX MEDLINE=91025552; PubMed=222811;
 RA Dull R.J., Uyeda C., Strosberg A.D., Nedwin G., Seilhamer J.J.;
 RT "Molecular cloning of cDNAs encoding bovine and human
 RT lactoperoxidase.";
 RL DNA Cell Biol. 9:489-509(1990).
 CC -1- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
 CC -1- COFACTOR: Binds 1 prothème IX and 1 iron(III) ion.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK AND SALIVARY GLAND.
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. XPO SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U39573; AAC50717.1; -;
 DR EMBL: M58151; AAA63213.1; -;
 DR PIR: J04935; J04935.
 DR HSRP: P05164; ICRP.
 DR GeneW: HGNC:6678; LPO.
 DR MIM: 150205; -;
 DR GO: GO:0016686; F:lactoperoxidase activity; NAS.
 DR GO: GO:0006804; P:peroxidase reaction; NAS.
 DR InterPro: IPR002007; Anim_peroxidase.
 DR InterPro: IPR002016; Peroxidase.
 DR Pfam: PF03096; An_peroxidase; 1.
 DR PRINTS: PR00457; ANPEROXIDASE.
 DR PROSITE: PS00435; PEROXIDASE_1; FALSE NEG.
 DR PROSITE: PS00436; PEROXIDASE_2; FALSE NEG.
 DR PROSITE: PS50292; PEROXIDASE_3; 1.
 DR Oxidoreductase; Peroxidase; Iron; Heme; Glycoprotein; Milk; Signal.
 KW Antidiabetic.
 KM
 FT SIGNAL 1 26 POTENTIAL.
 FT PROPEP 27 100
 FT CHAIN 101 712 LACTOPEROXIDASE.
 FT ACT_SITE 226 226 DISTAL HISTIDINE (POTENTIAL).
 FT ACT_SITE 372 372 DISTAL ARGININE (POTENTIAL).
 FT BINDING 225 225 HEME (COVALENT) (BY SIMILARITY).
 FT BINDING 375 375 HEME (COVALENT) (BY SIMILARITY).
 FT METIL 468 468 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 421 421 V -> M (IN REF. 2).
 SQ SEQUENCE 712 AA; 80287 MW; BC7AA410997198C0 CRC64;
 Query Match 72.0%; Score 59; DB 1; Length 712;
 Best Local Similarity 75.0%; Pred. No. 0.01;

Matches 12; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 TORTRNGFRVPLARE 16
| | | | | | | | | |
179 TPGRTRNGFRVPLARE 194

Db

RESULT 3
PERM_MOUSE STANDARD; PRT; 718 AA.
ID PERM_MOUSE
AC P11247;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myeloperoxidase precursor (EC 1.11.1.7) (MPO).
GN MPO.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H;
RA MEDLINE=89345187; PubMed=2548170;
RA Venturielli D., Shireat N., Gempferlein I., Bittenbender S.,
RA Rovera G.;
RT "Nucleotide sequence of cDNA for murine myeloperoxidase.",
RL Nucleic Acids Res. 17:5852-5852(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=90016884; PubMed=2552419;
RA Venturielli D., Bittenbender S., Rovera G.;
RT "Sequence of the murine myeloperoxidase (MPO) gene.",
RL Nucleic Acids Res. 17:7987-7988(1989).
CC -1- FUNCTION: Part of the host defense system of polymorphonuclear
leukocytes. It is responsible for microbicidal activity against a
wide range of organisms. In the stimulated PMN, MPO catalyzes the
production of hypochlorous acids, primarily hypochlorous acid in
physiologic situations, and other toxic intermediates that greatly
enhance PMN microbicidal activity (By similarity).
CC -1- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
CC -1- CATALYTIC ACTIVITY: Cl(-) + H(2)O(2) = HOCl + 2 H(2)O.
CC -1- COFACTOR: Binds 1 prochrome IX covalently, 1 iron(III) ion and 1
calcium ion per subunit (By similarity).
CC -1- SUBUNIT: Tetramer of two light chains and two heavy chains.
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. XPO SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15313; CA33373.1; -;
DR EMBL: X15378; CA33439.1; -;
DR PIR: S06068; S06068.
DR HSSP: P05164; 1CXP.
DR MGD: MGI:97137; MPO.
DR InterPro: IPR002007; Anim_peroxidase.
DR InterPro: IPR002016; Peroxidase.
DR Pfam: PF03098; An_peroxidase; 1.
DR PRINTS: PR00457; ANPEROXIDASE.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PROSITE: PS00436; PEROXIDASE_2; FALSE_NEG.
DR PROSITE: PS50292; PEROXIDASE_3; 1.
KM Oxidoreductase; peroxidase; Iron; Heme; Calcium-binding; Glycoprotein;
KW Signal; Lysosome.
FT SIGNAL 1 15
FT PROPEP 16 138 POTENTIAL.
FT CHAIN 139 718 MYELOPEROXIDASE.

FT CHAIN 139 252 MYELOPEROXIDASE LIGHT CHAIN.
FT CHAIN 253 718 MYELOPEROXIDASE HEAVY CHAIN.
FT ACT_SITE 235 235 DISTAL HISTIDINE (POTENTIAL).
FT ACT_SITE 379 379 DISTAL ARGININE (POTENTIAL).
FT METAL 476 476 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 711 711 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 718 AA; 81070 MW; D734401F8B4011B CRC64;

Query Match 62.24; Score 51; DB 1; Length 718;
Best Local Similarity 81.84; Pred. No. 0.27;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 6 RRGFRVPLARE 16
| | | | | | | | | |
193 RRGFRVPLAREQ 203

Db

RESULT 4
PERM_MOUSE STANDARD; PRT; 716 AA.
ID PERM_MOUSE
AC P49250; O61798;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Eosinophil peroxidase precursor (EC 1.11.1.7) (EPO).
GN EPX OR EPPER.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA Ohnori J., Itoh H., Tomita M., Nawa Y.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Bone marrow;
RX MEDLINE=96369651; PubMed=8773591;
RA Horton M.A., Larson K.A., Lee J.J., Lee N.A.;
RT "Cloning of the murine eosinophil peroxidase gene (mEPO):
RT characterization of a conserved subgroup of mammalian hematopoietic
peroxidases.",
RL J. Leukoc. Biol. 60:285-294(1996).
CC -1- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
CC -1- COFACTOR: Binds 1 prochrome IX and 1 iron(III) ion.
CC -1- SUBUNIT: Tetramer Of Two Light Chains And Two Heavy Chains (By
Similarity).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC GRANULES OF EOSINOPHILS.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. XPO SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D78353; BA11370.1; -;
DR EMBL: L79793; AAB4403.1; ALT_INIT.
DR HSSP: P05164; 1CXP.
DR MGD: MGI:107569; EPX.
DR InterPro: IPR002007; Anim_peroxidase.
DR InterPro: IPR002016; Peroxidase.
DR Pfam: PF03098; An_peroxidase; 1.
DR PRINTS: PR00457; ANPEROXIDASE.
DR PROSITE: PS00435; PEROXIDASE_1; 1.

RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagers K., Lacroix C., Maclean J., Moulé S.,
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrett B.G.,
RA "Massive gene decay in the leprosy bacillus.",
RA Nature 409:1007-1011(2001).
RL [3]
RP SEQUENCE OF 276-619 FROM N.A.
RX MEDLINE=89079289; PubMed=491836;
RA Garsia R.J., Helligvist L., Booth R.J., Radford A.J., Britton W.J.,
RA Asbury L., Trent R.J., Basten A.,
RA "Homology of the 70-kilodalton antigen from Mycobacterium leprae and
RA Mycobacterium bovis with the Mycobacterium tuberculosis 71-kilodalton
RA antigen and with the conserved heat shock protein 70 of eucaryotes.",
RL Infect. Immun. 57:204-212(1989).
RL [4]
RP PHOSPHORYLATION.
RX MEDLINE=98422335; PubMed=9748655;
RA Peake P., Winter N., Britton W.,
RA "Phosphorylation of Mycobacterium leprae heat-shock 70 protein at
RA threonine 175 alters its substrate binding characteristics.",
RL Biochim. Biophys. Acta 1387:387-394(1998).
RL [1-] FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).
CC [1-] INDUCTION: By stress conditions e.g. heat shock.
CC [1-] SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M95576; AAA25362.1; -;
DR EMBL; AL583925; CAC32013.1; -;
DR PIR; B87221; B87221.
DR HSSP; P08109; ICKR.
DR PhosSite; P19993; -;
DR Leptoma; ML2496; -;
DR HAMAP; MF_00332; -; 1.
DR InterPro; IPR01023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PRODOM; PD000089; Hsp70; 1.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00328; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DR KW Chaperone; ATP-binding; Heat shock; Phosphorylation;
KW Complete proteome.
FT INIT MET 0
FT MOD RES 174 174
FT CONFLICT 369 369
FT COMFLICT 607 618
FT FT
SQ SEQUENCE 619 AA; 66547 MW; 9490800A23A96612 CRC64;
Query Match 51.2%; Score 42; DB 1; Length 619;
Best Local Similarity 72.7%; Pred. No. 9.3;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

AC P11678;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Eosinophil peroxidase precursor (EC 1.11.1.7) (EPO).
GN EPO OR EPER OR EPO OR EPP.
OS Homo sapiens (Human).
OC Eukaryota; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISUE=Blood;
RX MEDLINE=89380315; PubMed=2550461;
RA Sakemaki K., Tomonaga M., Tsukui K., Nagata S.,
RA "Molecular cloning and characterization of a chromosomal gene for
RA human eosinophil peroxidase.",
RL J. Biol. Chem. 264:16828-16836(1989).
RN [2]
RP SEQUENCE OF 13-715 FROM N.A., AND SEQUENCE OF 140-187 AND 251-288.
RC TISUE=Blood;
RX MEDLINE=89235593; PubMed=2541222;
RA Ten R.M., Pease J.R., McKean D.J., Bell M.P., Gleich G.J.,
RA "Molecular cloning of the human eosinophil peroxidase. Evidence for
RA the existence of a peroxidase multigene family.",
RL J. Exp. Med. 165:1757-1769(1989).
RN [3]
RP COVALENT HEME ATTACHMENT, AND PARTIAL SEQUENCE.
RC TISUE=Blood;
RX MEDLINE=99287891; PubMed=10358043;
RA Oxyvig C., Thomsen A.R., Overgaard M.T., Sorensen E.S., Hoefjup P.,
RA Bjerrum M.J., Gleich G.J., Sottirup-Jensen L.,
RA "Biochemical evidence for heme linkage through esters with Asp-93 and
RA Glu-241 in human eosinophil peroxidase. The ester with Asp-93 is only
RA partially formed in vivo.",
RL J. Biol. Chem. 274:16953-16958(1999).
RN [4]
RP VARIANT EPO HIS-286.
RX MEDLINE=95107984; PubMed=7809065;
RA Romano M., Patriarca P., Melo C., Baralle F.E., Dri P.,
RA "Hereditary eosinophil peroxidase deficiency: immunochemical and
RA spectroscopic studies and evidence for a compound heterozygosity of
RA the defect.",
RL Proc. Natl. Acad. Sci. U.S.A. 91:12496-12500(1994).
CC [1-] CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
CC [1-] COFACTOR: Binds 1 prothemin IX and 1 iron(III) ion.
CC [1-] SUBUNIT: Tetramer of two light chains and two heavy chains.
CC [1-] SUBCELLULAR LOCATION: CYTOPLASMIC GRANULES OF EOSINOPHILS.
CC [1-] DISEASE: Defects in EPP are the cause of eosinophil peroxidase
CC deficiency (EPP) [MIM:261500]. EPP is an autosomal recessive
CC defect where anomalous eosinophils are characterized by nuclear
CC hypersegmentation, hypogranulation, and negative peroxidase and
CC phospholipid staining.
CC [1-] SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. XPO SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29913; AAA58458.1; -;
DR EMBL; M29904; AAA58458.1; JOINED.
DR EMBL; M29805; AAA58458.1; JOINED.
DR EMBL; M29806; AAA58458.1; JOINED.
DR EMBL; M29907; AAA58458.1; JOINED.
DR EMBL; M29908; AAA58458.1; JOINED.
DR EMBL; M29909; AAA58458.1; JOINED.
DR EMBL; M29910; AAA58458.1; JOINED.
DR EMBL; M29911; AAA58458.1; JOINED.
DR EMBL; M29912; AAA58458.1; JOINED.

```

DR EMBL: X14346; CAA32530.1; -.
DR PIR: A34408; A34408.
DR HSRF; P05164; ICKP.
DR Genew; HGNC:3423; EPX.
DR MIM; 131399; -.
DR InterPro; IPR002007; Anim.peroxidase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF01098; An.peroxidase; 1.
DR PRINTS; PR00457; ANPEROXIDASE.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; FALSE_NEG.
DR PROSITE; PS50292; PEROXIDASE_3; 1.
DR Oxidoreductase; peroxidase; Iron; Heme; Glycoprotein; Signal;
KW Disease mutation.
FT SIGNAL 1 17
FT PROPEP 18 139
FT CHAIN 140 250
FT CHAIN 251 715
FT BINDING 232 232
FT FT
FT BINDING 380 380
FT ACT SITE 233 233
FT ACT SITE 377 377
FT METAL 474 474
FT CARBOHD 52 52
FT CARBOHD 113 113
FT CARBOHD 327 327
FT CARBOHD 363 363
FT CARBOHD 700 700
FT CARBOHD 708 708
FT VARIANT 286 286
FT FT
FT CONFLICT 13 18
FT CONFLICT 21 21
FT CONFLICT 113 113
FT CONFLICT 163 163
FT CONFLICT 645 660
SQ SEQUENCE 715 AA; 81040 MW; CEA4E689A6C46374 CRC64;
Query Match 51.2%; Score 42; DB 1; Length 715;
Best Local Similarity 53.3%; Pred. No. 11;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 TORTRNGFRVPLAR 15
Db 186 TFSRRNGFLPLVR 200

```

```

RA Brown N.G.;
RA Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
RL (3)
RC SEQUENCE FROM N.A.
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A.,
RA Areujo R., Aparicio A., Barrell B.G., Badcock K., Bentes V.,
RA Borestein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo F., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kaiman S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirzaiati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schaefer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Sidic A.M., Tettelein H.,
RA Urrestarazu U.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wandut R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL Nature 387:103-105(1997).
CC -1- FUNCTION: HPI1/ADA1 AND SPT20/ADA5 MAY RECRUIT TATA BINDING
CC PROTEIN (TBP) AND POSSIBLY OTHER BASAL FACTORS TO BIND TO THE TATA
CC BOX WHILE ADA2, ADA3 AND GCN5 FUNCTION TO ACETYLATE NUCLEOSOMES
CC -1- SUBUNIT: PART OF THE ADA/GCN5 COMPLEX THAT CONSISTS OF HPI1/ADA1,
CC ADA2, ADA3, SPT20/ADA5 AND GCN5.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U76735; AAB58359.1; -.
DR EMBL: U41324; AAB4983.1; -.
DR EMBL: Z67551; CAA91590.1; -.
DR EMBL: Z73610; CAA97979.1; -.
DR PIR: S61010; S61010.
DR TRANSPAC; T04327; -.
DR SGD; S0006175; HPI1.
DR GO; GO:0000124; C:SAGA complex; IDA.
DR GO; GO:0003712; P:transcription cofactor activity; IMP.
DR GO; GO:0006366; P:transcription from Pol II promoter; IMP.
KW Transcription regulation; Nuclear protein.
FT CONFLICT 19 19
FT CONFLICT 19 19
SQ SEQUENCE 488 AA; 54466 MW; B3ACF3C610A5541 CRC64;
Query Match 50.0%; Score 41; DB 1; Length 488;
Best Local Similarity 80.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 7 NGFRVPLARE 16
Db 253 NGFRVPLASE 262

```

```

RESULT 9
HPI1_YEAST STANDARD; PRT; 488 AA.
AC Q12060; Q00039; Q02813;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transcriptional coactivator HPI1/ADA1.
GN HPI1 OR ADA1 OR SUP110 OR YPL254W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BK1-7A;
RX MEDLINE=97299672; PubMed=9154821;
RA Horituchi J., Silverman N., Pina R., Marcus G.A., Guarente L.;
RA "ADA1, a novel component of the ADA/GCN5 complex, has broader effects
RT than GCN5, ADA2, or ADA3."
RL Mol. Cell. Biol. 17:3220-3228(1997).
RN [2]
RP SEQUENCE FROM N.A.

```


OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OK NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=2196065; PubMed=12000796;
 RA Robertek S., Samesch I.E.;
 RT "Targets of ATM/ATR regulation during plant senescence and pathogen
 defense.",
 RT Genes Dev. 16:1139-1149(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Ramsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Romling C.M., Koo H.L.,
 RA Motilal K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
 RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
 RA Niernm W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
 RA Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana.",
 RT Nature 402:761-768(1999).
 RN [3]
 RP FUNCTION, AND INDUCTION.
 RX MEDLINE=2186469; PubMed=11875555;
 RA Asai T., Tena G., Plotnikova J., Willmann M.R., Chiu W.L.,
 RA Gomez-Gomez L., Bolter T., Ausubel F.M., Sheen J.;
 RT "MAP kinase signalling cascade in Arabidopsis innate immunity.",
 RT Nature 415:977-983(2002).
 CC -1- FUNCTION: Involved in innate immune response of plants.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- INDUCTION: Highly induced by WRKY22 or WRKY29 and by WRKY6 in
 CC senescent leaves. Also induced 30 minutes after flagellin
 CC treatment.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -1- SIMILARITY: Contains 3 leucine-rich (LRR) repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF46619; AAL92103.1; -;
 CC EMBL; AC002392; AAD12037.1; -;
 CC PIR; T00540; T00540.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR007090; LRR plant.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser_Thr_kinase.
 DR Pfam; PF00560; LRR_3.
 DR Pfam; PF00069; Kinase_1.
 DR ProDom; PD000001; Prot_kinase_1.
 DR ProSite; PS00107; PROTEIN KINASE ATP; 1.
 DR ProSite; PS00011; PROTEIN KINASE DOM; 1.
 DR ProSite; PS00108; PROTEIN KINASE_SF; 1.
 KW Plant defense; Transferrase; Serine/threonine-protein kinase; Receptor;
 KW ATP-binding; Transmembrane; Repeat; Leucine-rich repeat; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 876
 FT SENESCENCE-INDUCED RECEPTOR-LIKE
 FT SERINE/THREONINE KINASE.
 FT DOMAIN 25 517
 FT TRANSMEM 518 538
 FT DOMAIN 539 876
 FT REPEAT 415 438
 FT LRR 1.

FT REPEAT 439 462 LRR 2.
 FT REPEAT 463 483 LRR 3.
 FT DOMAIN 574 847 PROTEIN KINASE.
 FT NP_BIND 580 588 ATP (BY SIMILARITY).
 FT ACT_SITE 697 697 ATP (BY SIMILARITY).
 FT BINDING 601 601 ATP (BY SIMILARITY).
 SQ SEQUENCE 876 AA; 98711 MW; EEEIC13EE35BC786 CRC64;
 Query Match 50.0%; Score 41; DB 1; Length 876;
 Best Local Similarity 46.2%; Pred. No. 20;
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 QY 1 TORIKINGRVRPL 13
 DB 510 TKKKNKGVIPL 522
 RESULT 10
 EXD1_DROME STANDARD; PRT; 320 AA.
 AC Q9VEB3; Q95U84;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Brix domain containing protein 1 homolog.
 DE CG7993.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkelley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Baileson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Buttle K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C.M., Gabrielian A.B., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaislali M., Kalush P., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McInneson D.,
 RA Merkulov G., Mishina N.V., Moadarry C., Morris U., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheel F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stralton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasarman D.A., Weisscock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhu G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RT Science 287:2185-2195(2000).

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan X.H.,
RA Rubin G.M., Celinker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (by similarity).
CC -1- SIMILARITY: Contains 1 Brix domain.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE003721; AAF5514.1; ALT_SEQ.
DR EMBL; AY058248; ALU13477.1; -.
DR FLYBase; FBgn0038585; CG7993.
DR InterPro; IPR007109; Brix.
DR Pfam; PF04427; Brix; 1.
KM Hypothetical protein; Nuclear protein.
FT DOMAIN 30 243 Brix.
FT CONFLICT 188 243 I -> T (IN REF. 1).
SQ SEQUENCE 320 AA; 36509 MW; EE98936DD68B3703 CRC64;

Query Match 47.6%; Score 39; DB 1; Length 320;
Best Local Similarity 69.2%; Pred. No. 16;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 KTRNGRPVPLA 16
DB 10 KTRGKVTLLARE 22

RESULT 11
VP19_VZVD STANDARD; PRT; 483 AA.
AC P09276;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Capsid assembly and DNA maturation protein (Capsid protein VP19C).
GN 20
OS Varicella-zoster virus (strain Dumas) (VZV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=10338;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=86306657; PubMed=1018124;
RA Davison A.J., Scott J.B.;
RT "The complete DNA sequence of varicella-zoster virus.";
RL J. Gen. Virol. 67:1759-1816(1986).
CC -1- FUNCTION: COMPONENT OF THE BASAL LAYER IN WHICH THE CAPSIDS ARE
CC EMBEDDED. BINDS DNA (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP19C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04370; CAA27903.1; -.

```

```

DR PIR; B27343; WZBR20.
DR InterPro; IPR004999; Herpes_VP19C.
DR Pfam; PF03327; Herpes_VP19C; 1.
DR Capsid assembly; Coat protein; DNA-binding.
SQ SEQUENCE 483 AA; 53977 MW; A584CF30689FP91 CRC64;

Query Match 46.3%; Score 38; DB 1; Length 483;
Best Local Similarity 57.1%; Pred. No. 37;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 TORKTRGPRVPLA 14
DB 328 TORRQREGVRLHA 341

RESULT 12
IDVc_HAEIN STANDARD; PRT; 491 AA.
AC P44872;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ketol-acid reductoisomerase (EC 1.1.1.86) (Acetohydroxy-acid
DE isomerase) (Alpha-keto-beta-hydroxyacyl reductoisomerase).
GN IIVC OR H10682.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=1727;
RN 11
RP SEQUENCE FROM N.A.
RX STRAIN=Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RX Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utecherback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: (R)-2,3-dihydroxy-3-methylbutanoate + NADP(+)
CC = (S)-2-hydroxy-2-methyl-3-oxobutanoate + NADPH.
CC -1- PATHWAY: Valine and isoleucine biosynthesis; second step.
CC -1- SIMILARITY: Belongs to the ketol-acid reductoisomerase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32751; AAC23342.1; -.
DR PIR; B64086; B64086.
DR TIGR; H10682; -.
DR HAMAP; MF_00435; -.
DR InterPro; IPR005056; ACh_Isomrcdase.
DR Pfam; PF01450; IIVC; 1.
DR TIGRFAMs; TIGR00465; IIVC; 1.
KW Oxidoreductase; Branched-chain amino acid biosynthesis; NADP;
KW Complete proteome.
FT INIT_MET 0 BY SIMILARITY.
FT ACT_SITE 131 131 POTENTIAL.
SQ SEQUENCE 491 AA; 54108 MW; 049C20AD5C607076 CRC64;

Query Match 46.3%; Score 38; DB 1; Length 491;
Best Local Similarity 70.0%; Pred. No. 38;

```

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 ORKTRNGFRV 11
 |||
 Db 79 ORATENGFKV 88

RESULT 13
 ILVC_PASMU STANDARD; PRT; 491 AA.

AC O9CLF1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Ketol-acid reductoisomerase (EC 1.1.1.86) (aceto-hydroxy-acid isomerase)
 DE isomerase (Alpha-keto-beta-hydroxy)acyl reductoisomerase).
 GN ILVC OR PM1284.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RT May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
 CC -1- CATALYTIC ACTIVITY: (R)-2,3-dihydroxy-3-methylbutanoate + NADP(+) = (S)-2-hydroxy-2-methyl-3-oxobutanoate + NADPH.
 CC -1- PATHWAY: Valine and isoleucine biosynthesis; second step.
 CC -1- SIMILARITY: Belongs to the ketol-acid reductoisomerase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE006167; AK03368.1; -.
 DR HAMAP; MF_004357; 1.
 DR InterPro; IPR000506; ACh_isomrctse.
 DR Pfam; PF01450; ILVC; 1.
 DR TIGRFAMs; TIGR00465; ILVC; 1.
 KW Oxidoreductase; Branched-chain amino acid biosynthesis; NADP;
 KW Complete proteome.
 KW ACT SITE 132
 FT ACT SITE 132 POTENTIAL.
 SQ SPROUTER 491 AA; 54204 MW; 9442457D56F2039B CRC64;

Query Match 46.3%; Score 38; DB 1; Length 491;
 Best Local Similarity 70.0%; Pred. No. 38;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 ORKTRNGFRV 11
 |||
 Db 80 ORATENGFKV 89

RESULT 14
 SYL_FERPE STANDARD; PRT; 897 AA.

AC Q9XDB4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Isoleucyl-tRNA synthetase (EC 6.1.1.5) (Isoleucine--tRNA ligase) (IleRS).
 DE ILRS.
 GN ILES.
 OS Ferriidobacterium pennivorans.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae;
 OC Ferriidobacterium.

OX NCBI_TaxID=93466;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 9078;
 RX MEDLINE=99240435; PubMed=10224005;
 RA Bertoldo C., Dufferer F., Jorgensen P.L., Antaramkian G.;
 RT "Pullulanase type I from Ferriidobacterium pennivorans vens: cloning, characterization, and expression of the gene and biochemical
 RT characterization of the recombinant enzyme";
 RT Appl. Environ. Microbiol. 65:2084-2091 (1999).
 RL Appl. Environ. Microbiol. 65:2084-2091 (1999).
 CC -1- CATALYTIC ACTIVITY: ATP + L-Isoleucine + tRNA(Ile) = AMP +
 CC -1- SUBSTRATE: Binds 1 zinc ion per subunit (By similarity).
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF096662; AAD30388.1; -.
 DR HSSP; P41972; IFFY.
 DR InterPro; IPR002300; tRNA-synt_1a.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR InterPro; IPR002301; tRNA-synt_1le.
 DR Pfam; PF00133; tRNA-synt_1; 1.
 DR PRINTS; PR00984; TRNASYNTH1B.
 DR TIGRFAMs; TIGR00392; IleS; 1.
 DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Metal-binding; Zinc.
 KW SITE 42 52 "HIGH" REGION.
 FT SITE 575 579 "RMSKS" REGION.
 FT BINDING 578 578 ATP (BY SIMILARITY).
 SQ SEQUENCE 897 AA; 103511 MW; 729981A0C228343 CRC64;

Query Match 46.3%; Score 38; DB 1; Length 897;
 Best Local Similarity 58.3%; Pred. No. 69;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TORKTRNGFRV 12
 |||
 Db 63 TRYKTRNGFRV 74

RESULT 15
 HGPC_HAEIN STANDARD; PRT; 1066 AA.

AC Q9X432;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemoglobin and hemoglobin-haptoglobin binding protein C precursor.
 DE HGPC.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HI689 / Serotype B;
 RX MEDLINE=99270928; PubMed=1038475;
 RA Morton D.J., Whitby P.W., Scull T.L.;
 RT "Effect of multiple mutations in the hemoglobin- and haptoglobin-binding proteins, Hgpa, Hgpb, and Hgpc, of Haemophilus influenzae type b";
 RT Infect. Immun. 67:2729-2739 (1999).
 RL -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE

```

CC      HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED
CC      FOR HEME UPTAKE.
CC      -1- SUBCELLULAR LOCATION: Outer membrane.
CC      -1- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC      EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCA
CC      REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.
CC      ADDITION OR LOSS OF CCA REPEAT UNITS WOULD CHANGE THE READING
CC      FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC      REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
CC      WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
CC      -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC      HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR EMBL; AF094574; AAD33112.1; -.
CC      DR InterPro; IPR006970; PT.
CC      DR InterPro; IPR00531; Tonb_boxc.
CC      DR Pfam; PF04886; PT; 1.
CC      DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC      DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
CC      KW Outer membrane; Transport; Tonb box; Multigene family; Signal;
CC      Receptor; Repeat.
CC      FT SIGNAL 1 24
CC      FT CHAIN 25 1066
CC      FT DOMAIN 26 53
CC      FT REPEAT 26 29
CC      FT REPEAT 30 33
CC      FT REPEAT 34 37
CC      FT REPEAT 38 41
CC      FT REPEAT 42 45
CC      FT REPEAT 46 49
CC      FT REPEAT 50 53
CC      FT SITE 63 70
CC      FT SITE 1049 1066
CC      SQ SEQUENCE 1066 AA; 122593 MW; EFB88D5CE4247583 CRC64;
CC      TONB BOX.
CC      TONB C-TERMINAL BOX.
CC      Query Match 46.3%; Score 38; DB 1; Length 1066;
CC      Best Local Similarity 53.3%; Pred. No. 82;
CC      Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
CC      Oy 2 QKTRNGFRVPLARE 16
CC      Db 747 QKTRNGFRAPTSDE 761

```

Search completed: July 30, 2003, 16:24:32
 Job time : 6.2381 secs

Description
peroxidase [EC 1.10.3.1]
peroxidase [EC 1.10.3.1]
myeloperoxidase (MPO)
GAD65 family protein
acetylglutamate kinase
dnaK-type shock protein 70 kD heat shock p70
peroxidase [EC 1.10.3.1]
hypothetical protein
hypothetical protein
HFI protein - Yeast
hypothetical protein
serine/threonine-protein phosphatase
hypothetical protein
probable 30S ribosomal protein L19
antiproteinase inhibitor Na+/H+ antiporter
Mutator-like transposon
hypothetical protein
hypothetical protein
hypothetical protein
hydroxyindole O-methyltransferase
hypothetical protein
gene 20 protein
ketol-acid reductoisomerase
myo-inositol protein-kinase
acyl-CoA oxidase b

R;Kiser, C.; Caterina, J.; H

A:Reference: primary structure of bovine lactoperoxidase, a fourth member of a mammalian heme
 A:Accession: S16103; MUID:91266958; PMID:2050150
 A:Molecule type: protein
 A:Residues: 101-108, 'B',110-448, 'R',450-712 <CAL>
 C:Superfamily: myeloperoxidase; myeloperoxidase homology
 C:Keywords: chromoprotein; glycoprotein; heme; iron; metalloprotein; oxidoreductase
 F:1-11/Domain: signal sequence #status predicted <SIG>
 F:122-100/Domain: propeptide #status predicted <PRO>
 F:134-707/Domain: myeloperoxidase homology <MPX>
 F:1101-712/Product: peroxidase #status experimental <MAT>
 F:1105-212,322,358,449/Binding site: carboxylate (asn) (covalent) #status predicted
 F:132-145,446-256,250-274,354-365,573-630,671-696/Distal/ide bonds: #status predicted
 F:126,375/Binding site: hemediol (asp, glu) (covalent) #status predicted
 F:126/Active site: His (distal axial ligand) #status predicted
 F:372/Active site: Arg #status predicted
 F:468/Binding site: hemediol iron (His) (proximal axial ligand) #status predicted

Query Match 100.0%; Score 82; DB 1; Length 712;
 Best Local Similarity 100.0%; Pred. NO. 1.2e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TQRTKNGFRVPLARE 16
 |||||||
 DB 179 TQRTKNGFRVPLARE 194

RESULT 2
 JC4935
 peroxidase (EC 1.11.1.7) precursor - human
 N:Alternate names: lactoperoxidase (LPO); salivary peroxidase (SPO)
 C:Species: Homo sapiens (man)
 C:Date: 22-Oct-1996 #sequence_revision 01-Nov-1996 #text_change 01-Dec-2000
 C:Accession: JC4935, B58928
 R:Kistner, C.; Caerolina, U.; Engler, J.A.; Rahemtulla, B.; Rahemtulla, F.

[illegible]

F:234.382.383/Binding site: hemediol (Asp, Glu, Met) (covalent) #status predicted
F:235/Active site: His (distal axial ligand) #status predicted
F:379/Active site: Arg #status predicted
F:476/Binding site: hemediol iron (His) (proximal axial ligand) #status predicted

Query Match 62.2%; Score 51; DB 2; Length 718;
Best Local Similarity 81.8%; Pred. No. 0.48;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 193 RINGFVPLAR 203

RESULT 4
G82487
GSDER family protein VCA0217 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: G82487
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Claydon, R.A.; Gwin, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers, P
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406:477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: G82487
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-359 <HEI>
A:Cross-references: GB:AE004361; GB:AE003853; NID:G9657600; PIDN:AAF96129.1; GSPDB:GN001
C:Genetics:
A:Experimental source: serogroup O1, strain N16961; biotype El Tor
A:Gene: VCA0217
A:Map position: 2

Query Match 53.7%; Score 44; DB 2; Length 359;
Best Local Similarity 69.2%; Pred. No. 4.3;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 74 QKTFMGFRLPLA 86

RESULT 5
H82734
acetylglutamate kinase XF1001 [imported] - Xylella fastidiosa (strain 945c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: H82734
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A53328 below
A:Accession: H82734
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-438 <SIM>
A:Cross-references: GB:AE003938; GB:AE003849; NID:G9105935; PIDN:AAF83811.1; GSPDB:GN001
R:Stimpson, A.J.G.; Reinach, P.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Brons, M.R.S.; Benoit, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre, h
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; France, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitzjima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laizer
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, P
A:Authors: Martins, E.M.F.; Matsumura, A.Y.; Menck, C.F.M.; Mireca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

Query Match	51.2%	Score 42	DB 2	Length 620
Best Local Similarity	72.7%	Pred. No. 18		
Matches	8	Conservative	1	Mismatches
			2	Indels
			0	Gaps
0Y	2	QKRTKNGFRYP	12	
		:		
Db	527	QRETEKNGSRYP	537	

RESULT 8
 A34408
 peroxidase (EC 1.11.1.7), eosinophil, precursor - human
 CDate: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 24-Oct-2000
 CAccession: A34408; B60866; S04746
 R:Sakamaki, K.; Tomonaga, M.; Tsukui, K.; Nagata, S.
 J. Biol. Chem. 264, 16828-16836, 1989
 A>Title: Molecular cloning and characterization of a chromosomal gene for human eosinophil
 A:Accession: A34408
 A:Molecule type: DNA
 A:Residues: 1715 <SAK>
 A:Cross-references: GB:M29904; GB:M29905; GB:M29906; GB:M29907; GB:M29908; GB:M29909; GB:M29910; GB:M29911; GB:M29912; GB:M29913; GB:M29914; GB:M29915; GB:M29916; GB:M29917; GB:M29918; GB:M29919; GB:M29920; GB:M29921; GB:M29922; GB:M29923; GB:M29924; GB:M29925; GB:M29926; GB:M29927; GB:M29928; GB:M29929; GB:M29930; GB:M29931; GB:M29932; GB:M29933; GB:M29934; GB:M29935; GB:M29936; GB:M29937; GB:M29938; GB:M29939; GB:M29940; GB:M29941; GB:M29942; GB:M29943; GB:M29944; GB:M29945; GB:M29946; GB:M29947; GB:M29948; GB:M29949; GB:M29950; GB:M29951; GB:M29952; GB:M29953; GB:M29954; GB:M29955; GB:M29956; GB:M29957; GB:M29958; GB:M29959; GB:M29960; GB:M29961; GB:M29962; GB:M29963; GB:M29964; GB:M29965; GB:M29966; GB:M29967; GB:M29968; GB:M29969; GB:M29970; GB:M29971; GB:M29972; GB:M29973; GB:M29974; GB:M29975; GB:M29976; GB:M29977; GB:M29978; GB:M29979; GB:M29980; GB:M29981; GB:M29982; GB:M29983; GB:M29984; GB:M29985; GB:M29986; GB:M29987; GB:M29988; GB:M29989; GB:M29990; GB:M29991; GB:M29992; GB:M29993; GB:M29994; GB:M29995; GB:M29996; GB:M29997; GB:M29998; GB:M29999; GB:M30000; GB:M30001; GB:M30002; GB:M30003; GB:M30004; GB:M30005; GB:M30006; GB:M30007; GB:M30008; GB:M30009; GB:M30010; GB:M30011; GB:M30012; GB:M30013; GB:M30014; GB:M30015; GB:M30016; GB:M30017; GB:M30018; GB:M30019; GB:M30020; GB:M30021; GB:M30022; GB:M30023; GB:M30024; GB:M30025; GB:M30026; GB:M30027; GB:M30028; GB:M30029; GB:M30030; GB:M30031; GB:M30032; GB:M30033; GB:M30034; GB:M30035; GB:M30036; GB:M30037; GB:M30038; GB:M30039; GB:M30040; GB:M30041; GB:M30042; GB:M30043; GB:M30044; GB:M30045; GB:M30046; GB:M30047; GB:M30048; GB:M30049; GB:M30050; GB:M30051; GB:M30052; GB:M30053; GB:M30054; GB:M30055; GB:M30056; GB:M30057; GB:M30058; GB:M30059; GB:M30060; GB:M30061; GB:M30062; GB:M30063; GB:M30064; GB:M30065; GB:M30066; GB:M30067; GB:M30068; GB:M30069; GB:M30070; GB:M30071; GB:M30072; GB:M30073; GB:M30074; GB:M30075; GB:M30076; GB:M30077; GB:M30078; GB:M30079; GB:M30080; GB:M30081; GB:M30082; GB:M30083; GB:M30084; GB:M30085; GB:M30086; GB:M30087; GB:M30088; GB:M30089; GB:M30090; GB:M30091; GB:M30092; GB:M30093; GB:M30094; GB:M30095; GB:M30096; GB:M30097; GB:M30098; GB:M30099; GB:M30100; GB:M30101; GB:M30102; GB:M30103; GB:M30104; GB:M30105; GB:M30106; GB:M30107; GB:M30108; GB:M30109; GB:M30110; GB:M30111; GB:M30112; GB:M30113; GB:M30114; GB:M30115; GB:M30116; GB:M30117; GB:M30118; GB:M30119; GB:M30120; GB:M30121; GB:M30122; GB:M30123; GB:M30124; GB:M30125; GB:M30126; GB:M30127; GB:M30128; GB:M30129; GB:M30130; GB:M30131; GB:M30132; GB:M30133; GB:M30134; GB:M30135; GB:M30136; GB:M30137; GB:M30138; GB:M30139; GB:M30140; GB:M30141; GB:M30142; GB:M30143; GB:M30144; GB:M30145; GB:M30146; GB:M30147; GB:M30148; GB:M30149; GB:M30150; GB:M30151; GB:M30152; GB:M30153; GB:M30154; GB:M30155; GB:M30156; GB:M30157; GB:M30158; GB:M30159; GB:M30160; GB:M30161; GB:M30162; GB:M30163; GB:M30164; GB:M30165; GB:M30166; GB:M30167; GB:M30168; GB:M30169; GB:M30170; GB:M30171; GB:M30172; GB:M30173; GB:M30174; GB:M30175; GB:M30176; GB:M30177; GB:M30178; GB:M30179; GB:M30180; GB:M30181; GB:M30182; GB:M30183; GB:M30184; GB:M30185; GB:M30186; GB:M30187; GB:M30188; GB:M30189; GB:M30190; GB:M30191; GB:M30192; GB:M30193; GB:M30194; GB:M30195; GB:M30196; GB:M30197; GB:M30198; GB:M30199; GB:M30200; GB:M30201; GB:M30202; GB:M30203; GB:M30204; GB:M30205; GB:M30206; GB:M30207; GB:M30208; GB:M30209; GB:M30210; GB:M30211; GB:M30212; GB:M30213; GB:M30214; GB:M30215; GB:M30216; GB:M30217; GB:M30218; GB:M30219; GB:M30220; GB:M30221; GB:M30222; GB:M30223; GB:M30224; GB:M30225; GB:M30226; GB:M30227; GB:M30228; GB:M30229; GB:M30230; GB:M30231; GB:M30232; GB:M30233; GB:M30234; GB:M30235; GB:M30236; GB:M30237; GB:M30238; GB:M30239; GB:M30240; GB:M30241; GB:M30242; GB:M30243; GB:M30244; GB:M30245; GB:M30246; GB:M30247; GB:M30248; GB:M30249; GB:M30250; GB:M30251; GB:M30252; GB:M30253; GB:M30254; GB:M30255; GB:M30256; GB:M30257; GB:M30258; GB:M30259; GB:M30260; GB:M30261; GB:M30262; GB:M30263; GB:M30264; GB:M30265; GB:M30266; GB:M30267; GB:M30268; GB:M30269; GB:M30270; GB:M30271; GB:M30272; GB:M30273; GB:M30274; GB:M30275; GB:M30276; GB:M30277; GB:M30278; GB:M30279; GB:M30280; GB:M30281; GB:M30282; GB:M30283; GB:M30284; GB:M30285; GB:M30286; GB:M30287; GB:M30288; GB:M30289; GB:M30290; GB:M30291; GB:M30292; GB:M30293; GB:M30294; GB:M30295; GB:M30296; GB:M30297; GB:M30298; GB:M30299; GB:M30300; GB:M30301; GB:M30302; GB:M30303; GB:M30304; GB:M30305; GB:M30306; GB:M30307; GB:M30308; GB:M30309; GB:M30310; GB:M30311; GB:M30312; GB:M30313; GB:M30314; GB:M30315; GB:M30316; GB:M30317; GB:M30318; GB:M30319; GB:M30320; GB:M30321; GB:M30322; GB:M30323; GB:M30324; GB:M30325; GB:M30326; GB:M30327; GB:M30328; GB:M30329; GB:M30330; GB:M30331; GB:M30332; GB:M30333; GB:M30334; GB:M30335; GB:M30336; GB:M30337; GB:M30338; GB:M30339; GB:M30340; GB:M30341; GB:M30342; GB:M30343; GB:M30344; GB:M30345; GB:M30346; GB:M30347; GB:M30348; GB:M30349; GB:M30350; GB:M30351; GB:M30352; GB:M30353; GB:M30354; GB:M30355; GB:M30356; GB:M30357; GB:M30358; GB:M30359; GB:M30360; GB:M30361; GB:M30362; GB:M30363; GB:M30364; GB:M30365; GB:M30366; GB:M30367; GB:M30368; GB:M30369; GB:M30370; GB:M30371; GB:M30372; GB:M30373; GB:M30374; GB:M30375; GB:M30376; GB:M30377; GB:M30378; GB:M30379; GB:M30380; GB:M30381; GB:M30382; GB:M30383; GB

A:Cross-references: GB:AE003922; GB:AE003849; NID:g9105722; PIDN:MAF83640.1; GSPDB:GN001
 A:Experimental source: strain 945C
 R:Simpson, A.J.G.; Reinhard, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Canario, L.E.A.; Carraro, D.M.; Carver, H
 de-Nevo, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Author: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; From
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, U.P.; Krieger, J.E.; Kurame, E.E.; Laizy
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
 A:Author: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeira, D.A
 Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Author: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Teuhake, M.H.; Valada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF0830

Query Match 50.0%; Score 41; DB 2; Length 138;
 Best Local Similarity 64.3%; Pred. No. 5.5;
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 2 QKTRNGFRVPLAR 15
 DB 123 QKTRNGFRVPLAR 136

RESULT 10
 T33940
 Hypothetical protein R05H5.4 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
 C:Accession: T33940
 R:McMurry, A.
 submitted to the EMBL Data Library, March 1995
 A:Reference number: Z19821
 A:Accession: T33940
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-414 <M1>
 A:Cross-references: EMBL:Z48795; PIDN:CAA88727.1; GSPDB:GN00020; CESP:R05H5.4
 C:Experimental source: clone R05H5
 C:Genetics:
 A:Gene: CESP:R05H5.4
 A:Map position: 2
 A:Introns: 49/3; 125/3; 213/3; 327/3
 C:Superfamily: unassigned ser/thr or Tyr-specific protein kinases; protein kinase homolo

Query Match 50.0%; Score 41; DB 2; Length 414;
 Best Local Similarity 63.6%; Pred. No. 17;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 2 QKTRNGFRVPL 12
 DB 327 KKKTRAGYRLP 337

RESULT 11
 S61010
 Hfl1 protein - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: protein P1001; protein YPL254W
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 29-Oct-1999
 C:Accession: S61010; S65283; S65285
 R:Pohl, T.M.
 submitted to the EMBL Data Library, November 1995
 A:Reference number: S61010
 A:Accession: S61010
 A:Molecule type: DNA
 A:Residues: 1-488 <POH>
 A:Cross-references: EMBL:Z67751; NID:g1061234; PIDN:CAA91590.1; PID:g1061235
 R:Pohl, T.M.

submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64899
 A:Accession: S65283
 A:Molecule type: DNA
 A:Residues: 1-488 <POH>
 A:Cross-references: EMBL:Z73610; NID:g1370521; PIDN:CAA97979.1; PID:e246957; PID:g137052
 A:Experimental source: strain S288C (AB972)
 R:Mesenguy, F.; Dubois, R.; Vierendeels, F.; Schrems, B.
 submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64935
 A:Accession: S65285
 A:Molecule type: DNA
 A:Residues: 1-111 <MES>
 A:Cross-references: EMBL:Z73610; MIPS:YPL254W
 A:Experimental source: strain S288C (AB972)
 C:Genetics:
 A:Gene: SGD:HR1
 A:Cross-references: SGD:S0006175; MIPS:YPL254W
 A:Map position: 16L

Query Match 50.0%; Score 41; DB 2; Length 488;
 Best Local Similarity 80.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 7 NGFRVPLARE 16
 DB 253 NGFRVPLARE 262

RESULT 12
 B83404
 Hypothetical protein PA1941 (imported) - *Pseudomonas aeruginosa* (strain PA01)
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: B83404
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.D.; B
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lariby, K.; Lam,
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathc
 A:Reference number: A82950; NID:20437337; PMID:10984043
 A:Accession: B83404
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-631 <STO>
 A:Cross-references: GB:AE004620; GB:AE004091; NID:g9947929; PIDN:AG05329.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA1941

Query Match 50.0%; Score 41; DB 2; Length 631;
 Best Local Similarity 58.3%; Pred. No. 27;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 2 QKTRNGFRVPL 13
 DB 474 QKTRNGFRVPL 485

RESULT 13
 T00540
 serine/threonine-specific protein kinase (EC 2.7.1.1) T20R24.21 - *Arabidopsis thaliana*
 N:Alternate names: probable receptor-like protein kinase At2g19190
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Feb-2001
 C:Accession: T00540; F84573
 R:Roundley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
 submitted to the EMBL Data Library, July 1997
 A:Description: *Arabidopsis thaliana* chromosome II BAC T20R24 genomic sequence.
 A:Reference number: Z14167
 A:Accession: T00540
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA

A:Residues: 1-876 <ROU>
A:Cross-references: EMBL:AC002392; NID:G3176701; PIDDN:APD12037.1; PID:G3176723
A:Experimental source: Cultivar Columbia
A:RFLN: X.; Kaul, S.; Rounsley, S.D.; Shee, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
W.; Koo, H.; Moffit, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
Euseb, D.; Nierman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A:Reference number: AB4420; MUID:20083487; PMID:106117297
A:Accession: P84573
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-876 <STO>
A:Cross-references: GB:AE002093; NID:G3176723; PIDDN:APD12037.1; GSPDB:GN00139
C:Genetics:
A:Gene: T20K24.21; At2g19190
A:Map position: 2
A:Introns: 29/1; 211/2; 372/1; 419/2; 443/2; 467/2; 494/2; 548/1; 617/3; 660/1; 681/1; 7
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C:Keywords: phosphotransferase
F:574-849/Domains: protein kinase homology <KIN>

Query Match 50.0%; Score 41; DB 2; Length 876;
Best Local Similarity 46.2%; Pred. No. 38;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TORKRNGFRVPL 13
|::|::|::|
DB 510 TKKKRNGYITPL 522

C.A.; Li, J.H., Li, Y., Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marziani
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Smith, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Xu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A:Reference number: AB6141; MID:21016719; PMID:1130712
A:Accession: B86442
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1..235 <STO>
A:Cross-references: GB:AB005172; NID:g1136742; PIDN:AAG31323.1; GSDBB:GN00141
C:Genetics:
A:Map position: 1

Query Match 47.6%; Score 38; DB 2; Length 235;
Best Local Similarity 53.3%;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Oy 1 TORRTNGRRVPLAR 15
 | | | | |
Db 219 TYRKAHNGCRLEPRK 233

Search completed: July 30, 2003, 16:31.19
Job time : 12.1905 secs

RESULT 14
F71198
hypothetical protein PH1859 - Pyrococcus horikoshii
C/Species: Pyrococcus horikoshii
C/Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C/Accession: F711198
R/Kanazabayashi, Y.; Sawada, M.; Horikawa, H.; Haitaka, Y.; Hino, Y.; Yamamoto, S.; Sekita,
M.; Ohnuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchika,
DNA Rep. 5: 55-76, 1998
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A/Reference number: A71000; MWID:19834137; PMID:1679194
A/Accession: F711198
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 11-128 <RAW>
A/Cross-References: GB:AP000007; NID:g3236134; PIDN:BAA30981.1; PID:g3256298
A/Experimental source: strain OT3
A/Note: this accession replaces an interim accession for a sequence replaced by GenBank
C/Genetics:
A/Gene: PH1859
C/Superfamily: Pyrococcus horikoshii hypothetical protein PH1859

Query Match 47.6%; Score 39; DB 2; Length 128;
Best Local Similarity 40.0%; Pred. No. 12;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Cy 1 TORRTNGPVPVLR 15
 ::|::|::|::|::|::|
Db 11 SGRSKSGKGFYPLVR 25

RESULT 15
 B66442
 Probable 30S ribosomal protein S11 [imported] - *Arabidopsis thaliana*
 CSpecies: *Arabidopsis thaliana* (mouse-ear cress)
 CDate: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 CAccession: B66442
 RThelodopsis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Greasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000
 AAuthors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:23:41 ; Search time 11.2381 Seconds
(without alignments)
169.082 Million cell updates/sec

Title: US-09-787-070-7

Perfect score: 82

Sequence: 1 TORTRNGFRVPLAR 16

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cn2_6/ptodata/1/pubpaa/ECT_NEW_PUB.pep:*
3: /cn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cn2_6/ptodata/1/pubpaa/ECTS_PUBCOMB.pep:*
7: /cn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
14: /cn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	40	48.8	264	US-10-156-761-12999
2	40	48.8	1333	US-09-738-626-4048
3	39	47.6	1156	US-09-757-716-1
4	39	47.6	1325	US-09-864-761-35612
5	38	46.3	124	US-09-984-271-247
6	38	46.3	125	US-09-984-271-192
7	38	46.3	172	US-10-080-170-596
8	38	46.3	330	US-10-153-668-116
9	38	46.3	482	US-09-815-242-11079
10	38	46.3	502	US-09-738-626-4722
11	37	45.1	15	US-10-133-056-10
12	37	45.1	50	US-09-993-844-38
13	37	45.1	172	US-10-080-170-74
14	37	45.1	363	US-09-802-640-40
15	37	45.1	363	US-10-133-056-2

16	37	45.1	363	15	US-10-133-056-4	Sequence 4, Appl1
17	37	45.1	363	15	US-10-225-567A-330	Sequence 330, App
18	37	45.1	405	15	US-09-738-626-6717	Sequence 6717, Ap
19	37	45.1	420	9	US-09-946-142-2	Sequence 2, Appl1
20	37	45.1	421	15	US-10-156-761-13577	Sequence 13577, A
21	36	43.9	229	16	US-10-080-170-46	Sequence 46, Appl1
22	36	43.9	330	15	US-10-156-761-9078	Sequence 9078, Ap
23	35	42.7	48	9	US-09-925-299-1098	Sequence 1098, Ap
24	35	42.7	48	11	US-09-925-299-1098	Sequence 1098, Ap
25	35	42.7	84	11	US-09-764-691-3365	Sequence 3365, Ap
26	35	42.7	84	11	US-10-091-512-234	Sequence 234, Ap
27	35	42.7	107	9	US-09-864-761-42419	Sequence 42419, A
28	35	42.7	129	15	US-10-106-698-4874	Sequence 4874, Ap
29	35	42.7	129	9	US-09-828-366-26	Sequence 26, Appl1
30	35	42.7	135	15	US-10-028-072-276	Sequence 276, App
31	35	42.7	135	15	US-10-121-049-276	Sequence 276, App
32	35	42.7	135	15	US-10-123-904-276	Sequence 276, App
33	35	42.7	135	15	US-10-140-470-276	Sequence 276, App
34	35	42.7	135	15	US-10-175-746-276	Sequence 276, App
35	35	42.7	135	15	US-10-176-918-276	Sequence 276, App
36	35	42.7	135	15	US-10-137-865-276	Sequence 276, App
37	35	42.7	135	15	US-10-140-474-276	Sequence 276, App
38	35	42.7	135	15	US-10-142-431-276	Sequence 276, App
39	35	42.7	135	15	US-10-143-114-276	Sequence 276, App
40	35	42.7	135	15	US-10-140-002-276	Sequence 276, App
41	35	42.7	135	15	US-10-142-419-276	Sequence 276, App
42	35	42.7	135	15	US-10-123-262-276	Sequence 276, App
43	35	42.7	135	15	US-10-142-423-276	Sequence 276, App
44	35	42.7	135	15	US-10-121-050-276	Sequence 276, App
45	35	42.7	135	15	US-10-121-050-276	Sequence 276, App

ALIGNMENTS

RESULT 1
US-10-156-761-12999
; Sequence 12999, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12999
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis

US-10-156-761-12999

Query Match 48.8%; Score 40; DB 15; Length 264;
Best Local Similarity 64.3%; Pred. No. 31;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 TORTRNGFRVPLAR 15
|||
DB 141 TORTRNGFRVPLAR 154

RESULT 2
US-09-738-626-4048

Sequence 4048, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAMA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 4048
LENGTH: 1333
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4048

Query Match 48.8%; Score 40; DB 10; Length 1333;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 TRNGRVPLAR 16
DB 903 TRGVGVVVAAR 914

RESULT 3
US-09-757-716-1
Sequence 1, Application US/09757716
Patent No. US20010012515A1
GENERAL INFORMATION:
APPLICANT: Magna, Holly
APPLICANT: Schaefer, Paul
APPLICANT: Lawton, Michael
APPLICANT: Vocum, Sue
APPLICANT: Mitchell, Peter
APPLICANT: Hutchinson, Nancy
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHOHYDROLASE-2
FILE REFERENCE: PF-0420 US
CURRENT APPLICATION NUMBER: US/09/757,716
CURRENT FILING DATE: 2001-01-09
PRIOR APPLICATION NUMBER: US/08/996,083
PRIOR FILING DATE: 1997-12-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 1156
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte Clone No. US20010012515A1: 1388013
US-09-757-716-1

Query Match 47.6%; Score 39; DB 9; Length 1156;
Best Local Similarity 70.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 6 RNGRVPLAR 15
DB 856 RNRGRINLAK 865

RESULT 4
US-09-864-761-35612
Sequence 35612, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Accellra-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/532,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 35612
LENGTH: 1325
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000511.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN PERIL LIVER, SIGNAL = 0.98
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
OTHER INFORMATION: EST_HUMAN HIT: AM867076-1, EVALUATE 1.00e-103
OTHER INFORMATION: EST_HUMAN HIT: AM867076-1, EVALUATE 1.00e-103

OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUE 5.00e-63
US-09-864-761-35612

Query Match 47.6%; Score 39; DB 9; Length 1325;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 TRNGRVPRLP 16
DB 532 TRNGRVPRLP 543

RESULT 5
US-09-984-271-247
Sequence 247, Application US/09984271
Publication No. US20030040088A1
GENERAL INFORMATION:

APPLICANT: Rosen et al.
TITLE OF INVENTION: 71 Human Secreted Proteins
FILE REFERENCE: P2030P1
CURRENT APPLICATION NUMBER: US/09/984,271

PRIOR FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: 09/482,273
PRIOR FILING DATE: 2000-01-13

PRIOR APPLICATION NUMBER: PCT/US99/15849
PRIOR FILING DATE: 1998-07-14
PRIOR APPLICATION NUMBER: 60/092,921

PRIOR FILING DATE: 1998-07-15
PRIOR APPLICATION NUMBER: 60/092,922
PRIOR FILING DATE: 1998-07-15

PRIOR APPLICATION NUMBER: 60/092,956
PRIOR FILING DATE: 1998-07-15
NUMBER OF SEQ ID NOS: 267
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 247
LENGTH: 124
TYPE: PRT
ORGANISM: Homo sapiens
US-09-984-271-247

Query Match 46.3%; Score 38; DB 11; Length 124;
Best Local Similarity 77.8%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 KTRNGRVP 12
DB 112 KTRNGRVP 120

RESULT 6
US-09-984-271-192
Sequence 192, Application US/09984271
Publication No. US20030040088A1
GENERAL INFORMATION:

APPLICANT: Rosen et al.
TITLE OF INVENTION: 71 Human Secreted Proteins
FILE REFERENCE: P2030P1
CURRENT APPLICATION NUMBER: US/09/984,271

PRIOR FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: 09/482,273
PRIOR FILING DATE: 2000-01-13

PRIOR APPLICATION NUMBER: PCT/US99/15849
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: 60/092,921

PRIOR FILING DATE: 1998-07-15
PRIOR APPLICATION NUMBER: 60/092,922
PRIOR FILING DATE: 1998-07-15
PRIOR APPLICATION NUMBER: 60/092,956

PRIOR FILING DATE: 1998-07-15
NUMBER OF SEQ ID NOS: 267
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 192
LENGTH: 125

TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (125)
OTHER INFORMATION: Xaa equals stop translation
US-09-984-271-192

Query Match 46.3%; Score 38; DB 11; Length 125;
Best Local Similarity 77.8%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 KTRNGRVP 12
DB 112 KTRNGRVP 120

RESULT 7
US-10-080-170-596
Sequence 596, Application US/10080170
Publication No. US20030129601A1
GENERAL INFORMATION:

APPLICANT: COLE, S.T.
TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
TREATMENT OF MYCOBACTERIOSES

TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
TREATMENT OF MYCOBACTERIOSES
FILE REFERENCE: 03495.0218
CURRENT APPLICATION NUMBER: US/10/080,170

PRIOR FILING DATE: 2002-06-10
PRIOR APPLICATION NUMBER: 60/270,123
PRIOR FILING DATE: 2001-02-22

NUMBER OF SEQ ID NOS: 652
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 596
LENGTH: 172
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-10-080-170-596

Query Match 46.3%; Score 38; DB 16; Length 172;
Best Local Similarity 63.6%; Pred. No. 44;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 TRNGRVPRLP 15
DB 102 TRNGRVPRLP 112

RESULT 8
US-10-153-668-116
Sequence 116, Application US/10153668
Publication No. US20030092616A1
GENERAL INFORMATION:

APPLICANT: MATSUDA, Akio
APPLICANT: MURAMATSU, Shuji
APPLICANT: ISHIZAMA, Kenya
TITLE OF INVENTION: STAT6 Activating Gene

FILE REFERENCE: 1254-020P
CURRENT APPLICATION NUMBER: US/10/153,668
PRIOR FILING DATE: 2002-05-24

PRIOR APPLICATION NUMBER: US 60/293,172
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/316,031

PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,403
PRIOR FILING DATE: 2001-10-12

PRIOR APPLICATION NUMBER: JP 2001-157043
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: JP 2001-260681
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: JP 2001-313175
PRIOR FILING DATE: 2001-10-10

NUMBER OF SEQ ID NOS: 488
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 116
LENGTH: 330
TYPE: PRT
ORGANISM: Homo sapiens
US-10-153-668-116

Query Match
Best Local Similarity 46.3%; Score 38; DB 15; Length 330;
Best Local Similarity 87.5%; Pred. No. 88;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 ORKTRNGFV 9
DB 137 ORKTRNGFV 144

RESULT 9
US-09-815-242-11079
Sequence 11079, Application US/09815242
Patent No. US2002061569A1
GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl U.
APPLICANT: Zyckind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 11079
LENGTH: 492
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-815-242-11079

Query Match
Best Local Similarity 46.3%; Score 38; DB 9; Length 492;
Best Local Similarity 70.0%; Pred. No. 1,4e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 ORKTRNGFV 11
DB 80 ORKTRNGFV 89

RESULT 10
US-09-738-626-4722
Sequence 4722, Application US/09738626
Patent No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: YATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO: 4722
LENGTH: 502
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4722

Query Match
Best Local Similarity 46.3%; Score 38; DB 10; Length 502;
Best Local Similarity 53.3%; Pred. No. 1,4e+02;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 ORKTRNGFVPLARE 16
DB 454 QFNTENHLPVPLRE 468

RESULT 11
US-10-133-056-10
Sequence 10, Application US/10133056
Patent No. US2003008339A1
GENERAL INFORMATION:
APPLICANT: Tamura, Masaaki
TITLE OF INVENTION: Compositions and Methods of Treating Colorectal Polyps and Cancer
FILE REFERENCE: Vanderbilt Ref. No. US2003008339A1 V00112; Our Ref. No. US200300833
CURRENT APPLICATION NUMBER: US/10/133,056
CURRENT FILING DATE: 2002-04-26
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 10
LENGTH: 15
TYPE: PRT
ORGANISM: Mus musculus
US-10-133-056-10

Query Match
Best Local Similarity 45.1%; Score 37; DB 15; Length 15;
Best Local Similarity 58.3%; Pred. No. 4,8;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 ORKTRNGFVPL 13
DB 1 QOKLRVFRVPI 12

RESULT 12
US-09-993-844-38
Sequence 38, Application US/09993844
Patent No. US20020106739A1
GENERAL INFORMATION:
APPLICANT: Oakley, Robert H.
APPLICANT: Barak, Lawrence S.
APPLICANT: Laporte, Stephanie A.
APPLICANT: Caron, Marc G.
TITLE OF INVENTION: Modified G-Protein Coupled Receptors
FILE REFERENCE: 033072-026

```

; CURRENT APPLICATION NUMBER: US/09/993,844
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 60/245,772
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/260,363
; PRIOR FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-844-38

Query Match      45.1%; Score 37; DB 10; Length 50;
Best Local Similarity 58.3%; Pred. No. 17;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      2 ORKTRNGFRVPL 13
        |||:|:|:|:|:|
Db      13 QOKLRSVFRVPI 24

RESULT 13
US-10-080-170-74
; Sequence 74, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLB, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; FILE REFERENCE: 03495,0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 74
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-080-170-74

Query Match      45.1%; Score 37; DB 16; Length 172;
Best Local Similarity 63.6%; Pred. No. 66;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      5 TRNGFRVPLAR 15
        |||:|:|:|:|
Db      102 TRSGVDIPLAR 112

RESULT 14
US-09-802-640-40
; Sequence 40, Application US/09802640
; Publication No. US20030036057A1
; GENERAL INFORMATION:
; APPLICANT: Braun, Andreas
; APPLICANT: Bonnal Aruna
; APPLICANT: Kleyn Patrick
; TITLE OF INVENTION: GENES AND POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: 24736-2048
; CURRENT APPLICATION NUMBER: US/09/802,640
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapien
```

```

US-09-802-640-40

Query Match      45.1%; Score 37; DB 11; Length 363;
Best Local Similarity 58.3%; Pred. No. 1.5e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      2 ORKTRNGFRVPL 13
        |||:|:|:|:|
Db      326 QOKLRSVFRVPI 337

RESULT 15
US-10-133-056-2
; Sequence 2, Application US/10133056
; Publication No. US20030083339A1
; GENERAL INFORMATION:
; APPLICANT: Tanura, Masaaki
; TITLE OF INVENTION: Compositions and Methods of Treating Colorectal Polyps and Cancer
; FILE REFERENCE: Vanderbilt Ref. No. US20030083339A1 V00112; Our Ref. No. US200300833
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Version 3.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-133-056-2

Query Match      45.1%; Score 37; DB 15; Length 363;
Best Local Similarity 58.3%; Pred. No. 1.5e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      2 ORKTRNGFRVPL 13
        |||:|:|:|:|
Db      326 QOKLRSVFRVPI 337

Search completed: July 30, 2003, 16:35:14
Job time : 12.2381 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 30, 2003, 16:09:50 ; Search time 31.5238 Seconds
(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-7
Perfect score: 82
Sequence: 1 TORKTRNGFRVLIARR 16

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	82	100.0	16	21	AAV88342	Peptide derived fr
2	82	100.0	712	12	AAR12124	Bovine lactoperox
3	43	52.4	525	23	ABG05707	Novel human diagn
4	42	51.2	715	23	ABR69521	Human polypeptide
5	42	51.2	1394	23	ABR66310	Bifidobacterium 10
6	41	50.0	81	23	ABP02245	Human ORFX protein
7	41	50.0	876	23	ABR91807	Herbically activ
8	40	48.8	101	22	AAW06693	Human foetal prote
9	40	48.8	294	21	AAQ35223	Zea mays protein f

10	40	48.8	339	21	AAQ35222	Zea mays protein f
11	40	48.8	391	21	AAQ35221	Zea mays protein f
12	40	48.8	1333	22	AAQ90294	C glutathione prote
13	39	47.6	62	22	AAU14545	Novel human protei
14	39	47.6	95	22	ABG07345	Novel human diagn
15	39	47.6	243	22	ABG02544	Novel human diagn
16	39	47.6	243	22	ABG03056	Novel human diagn
17	39	47.6	515	21	AAV77952	A. thaliana enviro
18	39	47.6	659	21	ABG00999	Novel human diagn
19	39	47.6	687	22	ABG19134	Novel human diagn
20	39	47.6	1156	20	AAV23754	Human nucleotide p
21	39	47.6	1325	22	ABG49735	Human liver peptid
22	39	47.6	1325	22	ABR29725	Peptide #2376 enco
23	39	47.6	1325	22	ABR34897	Peptide #2403 enco
24	39	47.6	1325	22	ABR20314	Protein #2313 enco
25	39	47.6	1325	22	AAW55707	Human brain expres
26	39	47.6	1325	22	AAW68085	Human bone marrow
27	39	47.6	1325	22	AAW06445	Peptide #2327 enco
28	39	47.6	1325	23	ABG37612	Human peptide enco
29	39	47.6	1684	22	ABR12028	Human d01042K10.4
30	39	47.6	1684	22	AAW79579	Human protein SEQ
31	39	47.6	1723	39	AAW78595	Human protein SEQ
32	38	46.3	69	22	AAU58868	Protonibacterium
33	38	46.3	124	21	AAV87208	Human secreted pro
34	38	46.3	124	22	AAW06130	Human gene 58 enco
35	38	46.3	124	22	AAW06185	Human gene 58 enco
36	38	46.3	124	22	ABG33953	Human secreted pro
37	38	46.3	124	23	ABG34009	Human secreted pro
38	38	46.3	124	21	AAV87153	Human secreted pro
39	38	46.3	159	22	ABG16246	Novel human diagn
40	38	46.3	172	20	AAV04964	Mycobacterium spec
41	38	46.3	112	23	AAU05945	M. tuberculosis an
42	38	46.3	166	20	AAV04965	Mycobacterium spec
43	38	46.3	225	22	AAW39172	Human polypeptide
44	38	46.3	253	23	AAE14443	Human drug metabol
45	38	46.3	330	21	AAV91506	Human secreted pro

ALIGNMENTS

RESULT 1
ID AAV88342
AAV88342 standard, peptide, 16 AA.
XX
AC AAV88342;
XX
XX
DT 14-JUL-2000 (first entry)
XX
DB Peptide derived from an internal fragment of sheep whey.
XX
XX Peptide production; biological fluid; milk; whey; blood;
KW antibacterial peptide; lactoferrin; antiviral; antitumour activity.
XX
XX
OS Ovis sp.
XX
XX
PN WO200015655-A1.
XX
PD 23-MAR-2000.
XX
PF 15-SEP-1999; 99WO-EP07002.
XX
PR 15-SEP-1998; 98EP-0203107.
PR 08-JUN-1999; 99EP-0201815.
PA (NIZO-) NIZO FOOD RES.
XX
PI Visser S, Recio I;
XX
XX WPI, 2000-271377/23.
DR
PT Novel process for producing peptides with e.g. antimicrobial activity
PT From biological fluids such as milk, whey or blood comprises contacting

PT Fluid with chromatographic medium to adsorb peptide domain of interest
XX
PS Claim 14; Page 22; 41pp; English.
XX
CC This sequence represents a peptide derived from an internal fragment of
CC sheep whey. The peptide is an example of a peptide with antibacterial
CC activity that can be produced by the process of the invention. The
CC invention relates to a process for producing peptides from biological
CC fluids. The process comprises chromatography of the biological fluid in
CC situ hydrolysis of selectively bound peptides, washing to remove unbound
CC peptide, and elution of the peptides of interest. The process is used for
CC producing peptides from biological fluids, such as milk, whey or blood.
CC For example, the process can be used to produce antibacterial peptides
CC derived from lactoferrin, using cheese whey as a starting material. The
CC peptides obtained have preferably antimicrobial and/or antiviral and/or
CC antitumour activity. The process of the invention is relatively simple
CC and generally economically and technically more attractive than those
CC methods previously used. The method provides high yield peptides with a
CC selected activity of interest without the need for intermediate
CC purification of the precursor protein.
XX
SQ Sequence 16 AA;
Query Match 100.0%; Score 82; DB 21; Length 16;
Best Local Similarity 100.0%; Pred. No. 7.2e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TORKTRNGFRVPLARE 16
DB 1 TORKTRNGFRVPLARE 16
RESULT 2
AAR12124
ID AAR12124 standard; Protein; 712 AA.
XX
AC AAR12124;
XX
DT 02-AUG-1991 (first entry)
DE Bovine lactoperoxidase prepro enzyme.
XX
KW LPO; autoimmune disease; antimicrobial; crosslinking agent;
KW Immunoassay; preservative.
XX
OS Bos taurus.
XX
FH Key Location/Qualifiers
FT Peptide 1..26
FT /label= signal sequence
FT /note= "or 1..28"
FT Protein 101..712
FT /label= mature LPO
FT /note= "or 129..712"
XX
PN WO9106639-A.
PD 16-MAY-1991.
XX
PF 31-OCT-1990; 90WO-US06317.
XX
PR 03-NOV-1989; 89US-0431634.
XX
PA (IDEO-) IDEON CORP.
XX
PI Sellhammer UT, Dull TJ;
XX
DR MPI; 1991-164189/22.
XX
DR N-PSDB; AAQ11842.
PT Recombinant lacto-peroxidase and deoxyribonucleic acid encoding
PT it - used to treat auto-immune diseases, preserve prod. e.g.

PT cosmetics, and to crosslink ligands to proteins.
XX
PS Claim 1; Fig 1; 31pp; English.
XX
CC The sequence was deduced from a cDNA library prep. from a bovine
CC milk gland. The sequence comprises 712 amino acids including the
CC signal sequence and prosequence. Two putative signal peptidase
CC cleavage sites and two possible alternative termini of the cleaved
CC propeptide are indicated. The LPO protein is an antimicrobial
CC useful for the treatment of infectious diseases and for preserva-
CC tion of perishable goods such as foods, beverages, and cosmetics.
CC It is active against fungi, viruses, protozoa and bacteria, and is
CC also effective against malignancies, autoimmune disease and trans-
CC plant rejection. It can also be used as an agent for crosslinking
CC protein to ligands. See also AAR12562.
XX
SQ Sequence 712 AA;
Query Match 100.0%; Score 82; DB 12; Length 712;
Best Local Similarity 100.0%; Pred. No. 3.8e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TORKTRNGFRVPLARE 16
DB 179 TORKTRNGFRVPLARE 194
RESULT 3
ABG05707
ID ABG05707 standard; Protein; 525 AA.
XX
AC ABG05707;
XX
DT 13-FEB-2002 (first entry)
DE Novel human diagnostic protein #5698.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HUSE-) HUSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR MPI; 2001-639362/73.
XX
DR N-PSDB; AAS69894.
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 3606; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or

quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (ii) and its binding partners are useful in medical imaging of sites expressing (ii). (i) and (ii) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 525 AA;

Query Match 52.4%; Score 43; DB 22; Length 525;
Best Local Similarity 75.0%; Pred. No. 31;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TORSTRNGFRVP 12

Db 456 TORSTRNGFRVP 467

RESULT 4

ID ABP69521 standard; Protein; 715 AA.

AC ABP69521;

DT 20-JAN-2003 (first entry)

DE Human polypeptide SEQ ID NO 1568.

Human; genome mapping; gene therapy; food supplement; virus; fungus; cell-proliferative disorder; neurodegenerative disease; bacterial; Parkinson's disease; Alzheimer's disease; autoimmune disease; multiple sclerosis; diabetes; genetic disorder; wound; burn; infection; arthritis; cytotoxic; immunomodulator; nootropic; neuroprotective; antiparkinsonian; antidiabetic; immunosuppressive; dermatological; haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide; antiarthritic.

OS Homo sapiens.

PN WO200270539-A2.

PD 12-SEP-2002.

PE 05-MAR-2002; 2002MO-US05095.

PR 05-MAR-2001; 2001US-0799451.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F, Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M, Wehrman T, Wang D, Drmanac RT;

PI WPI; 2002-759812/82.

DR N-PSDB; AB211738.

New polynucleotides comprising sequences assembled from expressed sequence tags (ESTs), useful for treating cell-proliferative, neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet or coagulation disorders.

PS Claim 9; SEQ ID NO 1568; 1012pp + Sequence Listing; English.

The invention relates to an isolated polynucleotide (I) comprising a nucleotide sequence selected from any of 948 sequences (AB21119-AB212066) or their mature protein coding portion, active domain

coding protein or complementary sequences. The polynucleotides are useful for identifying expressed genes or for physical mapping of human genome. The encoded polypeptides (ABP6902-ABP6949) are useful as molecular weight markers, as a food supplement, for generating antibodies, in medical imaging, screening and diagnostic assays and for treating cell-proliferative disorders (cancer), neurodegenerative diseases (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid disorders, platelet or coagulation disorders, wound, burns, infection, ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal, parasitic), arthritis, etc.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 715 AA;

Query Match 51.2%; Score 42; DB 23; Length 715;
Best Local Similarity 53.3%; Pred. No. 65;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 TORSTRNGFRVP 15

Db 186 TPRSRRNGFLLPLVR 200

RESULT 5

ID ABP66310 standard; Protein; 1394 AA.

AC ABP66310;

DT 19-NOV-2002 (first entry)

DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:1054.

Bifidobacterium longum NCC2705; Bifidobacterium; bacterial; antidiarrhetic; antibacterial; inhibitor of Salmonella; detection; identification; lactic acid bacterium; diarrhoea; pathogenic bacteria; rotavirus; food composition; pharmaceutical composition.

OS Bifidobacterium longum.

PN EP1227152-A1.

PD 31-JUL-2002.

PE 30-JAN-2001; 2001EP-0102050.

PR 30-JAN-2001; 2001EP-0102050.

PA (NEST) SOC PROD NESTLE SA.

PI WPI; 2002-668397/72.

PI Novel polynucleotide comprising Bifidobacterium genome sequence useful as a probe or primer for detecting and/or identifying Bifidobacterium

PI longum in a biological sample.

PS Claim 3; SEQ ID 1054; 80pp; English.

The present invention describes a polynucleotide (I) comprising a sequence of a Bifidobacterium genome selected from the nucleotide sequences given in AB081842 and AB081843, or a sequence exhibiting at least 90% identity or which hybridises with the sequences given in AB081842 and AB081843. Also described is a polynucleotide (II) encoding a fusion protein, comprising a sequence selected from 1097 sequences given in ABP65258 to ABP6534 ligated in frame to a polynucleotide encoding a heterologous polypeptide. (I) has antidiarrhetic and antibacterial activities, and can be used as an inhibitor of Salmonella. (II) (which is a probe) is useful for the detection and/or identification of Bifidobacterium longum in a biological sample. A carrier containing the lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM I-2618)

CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

XX Sequence 876 AA;

Query Match 50.0%; Score 41; DB 23; Length 876;

Best Local Similarity 46.2%; Pred. No. 1.2e+02; Mismatches 5; Indels 0; Gaps 0;

Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 TORKTRNGFRVPL 13

DB 510 TKKONKNGYIPL 522

RESULT 8

AA06693 ID AA06693 standard; Protein; 101 AA.

AC AA06693;

DT 05-OCT-2001 (first entry)

DE Human foetal protein, SEQ ID NO: 424.

KM Human; foetal protein; cytotaxic; immunosuppressive; immunostimulant;

KW neurotrophic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;

KW gene therapy; antitense therapy; cancer; immune disorder;

KW growth disorder; osteoporosis; thrombolytic disorder;

KW nervous system disorder; inflammation.

OS Homo sapiens.

PN WO200155339-A2.

PD 02-AUG-2001.

PF 25-JAN-2001; 2001WO-US02723.

PR 25-JAN-2000; 2000US-0491404.

PR 15-SEP-2000; 2000US-0663870.

PR 06-NOV-2000; 2000US-0707351.

PA (HYSB-) HXSEQ INC.

PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Dermanac RA, Tang YT;

PI Liu C, Aaundi V, Zhou P, Werhman T;

DR MPI: 2001-465571/50.

DR N-PsDB; AAH94368.

PT Novel fetal proteins useful for the treatment and diagnosis of diseases

PT associated with dysfunction of the protein e.g. cancers, immune

PT disorders, growth disorders, thrombolytic disorders, nervous system

PT disorders and inflammation.

PS Claim 10; Page 332; 715pp; English.

XX The invention relates to novel foetal polypeptides encoded by

XX polynucleotides comprising one of 477 sequences fully defined in the

XX specification. The foetal polynucleotides and polypeptides are

XX useful in the treatment and diagnosis of diseases such as cancers,

XX immune disorders, growth disorders (e.g. osteoporosis), thrombolytic

XX disorders, nervous system disorders and inflammation. The present

XX sequence is a polypeptide encoded by a cDNA assembled using

XX an expressed sequence tag (EST) found to be expressed in human

XX foetal tissue cDNA libraries.

XX Sequence 101 AA;

XX Query Match 48.8%; Score 40; DB 22; Length 101;

Best Local Similarity 70.0%; Pred. No. 19;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QKTRNGFRV 11

DB 87 QKTRNGFOL 96

RESULT 9

AA035223 ID AA035223 standard; Protein; 294 AA.

AC AA035223;

DT 18-OCT-2000 (first entry)

DE Zea mays protein fragment SEQ ID NO: 42997.

KM Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence; corn.

OS Zea mays subsp. mays.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 23-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0129845.

PR 21-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135553.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139839.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147312.
PR 05-AUG-1999; 99US-0147320.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.

Query Match

Best Local Similarity 48.8%; Score 40; DB 21; Length 294;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TORRTNGRPVPIARE 16
Db 5 TTRKTRDPFAIKARE 20

RESULT 10
AAG35222
ID AAG35222 standard; Protein: 339 AA.
XX AAG35222;
AC
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 42996.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence; corn.
XX
XX Zea mays subsp. mays.
OS
XX
XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
BD
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134286.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134991.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135639.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139453.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142955.
PR 06-JUL-1999; 99US-0142990.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145819.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155466.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158222.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 48.8%; Score 40; DB 21; Length 339;
Best Local Similarity 56.2%; Pred. No. 68;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TORKTRNGFVFLARE 16
| | | | : | | | |
Db 50 TTRKTRDPFAIKKARE 65

RESULT 11
AAG35221
ID AAG35221 standard; Protein; 391 AA.
XX
AC AAG35221;
XX

DT 18-OCT-2000 (first entry)
XX Zea mays protein fragment SEQ ID NO: 42995.
DE Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX EPI033405-A2.
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135553.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143564.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144652.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145182.
PR 23-JUL-1999; 99US-0145185.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147182.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149928.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157173.
PR 06-OCT-1999; 99US-0157863.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 48.8%; Score 40; DB 21; Length 391;
Best local similarity 56.2%; Pred. No. 79;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 TORKRNQFRRPLARE 16
| | | | | : | : | | |
Db 102 TTRKTHDPFAIRKARE 117

RESULT 12
AAG90294
ID AAG90294 standard; Protein; 1333 AA.
XX AAG90294;
AC
XX
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 4048.
XX
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX

OS Corynebacterium glutamicum.
 XX EPI108790-A2.
 XX
 XX 20-JUN-2001.
 XX
 XX 18-DEC-2000; 2000EP-0127688.
 XX
 XX 16-DEC-1999; 99JP-0377484.
 XX PR 07-APR-2000; 2000JP-0159162.
 XX PR 03-AUG-2000; 2000JP-0280988.
 XX
 XX (KYOWA) KYOWA HAKKO KOSYO KK.
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H,
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 XX MPI: 2001-376931/40.
 XX N-PSDB; AAH65513.
 XX
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 XX Claim 17; SEQ ID NO: 4048; 246pp + Sequence Listing; English.
 XX
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 XX Sequence 1333 AA;
 SQ
 Query Match 48.8%; Score 40; DB 22; Length 1333;
 Best Local Similarity 66.7%; Pred. No. 2.9e+02;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 OY 5 TNGRGPVPLARR 16
 Db 903 TRQGVPRVVAAR 914
 RESULT 13
 AAU14545
 ID AAU14545 standard; Protein; 62 AA.
 XX
 XX AAU14545;
 XX
 XX 24-OCT-2001 (first entry)
 XX
 XX Human novel protein #416.
 XX
 XX Human; novel protein; Antianemic; osteopathic; antiinflammatory;
 KW immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
 KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antisthmatic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 KW tissue regeneration; immune disorder.
 XX
 XX Homo sapiens.
 OS
 XX MO200155437-A2.
 PN

XX
 XX 02-AUG-2001.
 PD
 XX 25-JAN-2001; 2001WO-US02623.
 XX PF
 XX 25-JAN-2000; 2000US-0491404.
 XX PR
 XX (HYSEQ-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Dymnac RT;
 PI
 XX MPI: 2001-451939/48.
 XX N-PSDB; AAS22850.
 XX
 XX Isolated polypeptides useful for treating anti-inflammatory diseases,
 PT nervous system disorders, and for regenerating bone and cartilage -
 PT
 XX
 XX Example 4; Page 866; 894pp; English.
 XX
 XX The invention relates to polynucleotides encoding novel human
 CC proteins or their active domains. The polypeptides, polynucleotides and
 CC antibodies raised against the polypeptides are used in a method of
 CC treatment of a mammal and prevention of disorders caused by the aberrant
 CC protein expression or activity. The polypeptides can be used as
 CC molecular weight markers, food supplements, and in antibody production.
 CC The polypeptides are used to identify compounds which bind to the
 CC polypeptides. Polynucleotides of the invention are used as probes and
 CC primers, for sequencing, for chromosome or gene mapping, in the
 CC production of recombinant proteins, and in generating anti-sense DNA or
 CC RNA and in gene therapy. Polypeptides of the invention can be used to
 CC target drugs to a tumour, in assays to determine biological activity, to
 CC raise antibodies/elicit an immune response, to determine quantitative
 CC protein levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,
 CC graft-versus-host disease, eczema, hemophilia, thrombosis,
 CC anti-inflammatory diseases, nervous system disorders, and infection.
 CC The present sequence represents a protein of the invention.
 XX
 XX Sequence 62 AA;
 SQ
 Query Match 47.6%; Score 39; DB 22; Length 62;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 OY 4 KTRNGRPV 12
 Db 32 ETRDGRFP 40
 RESULT 14
 ABG07345
 ID ABG07345 standard; Protein; 95 AA.
 XX
 XX ABG07345;
 XX
 XX 13-FEB-2002 (first entry)
 XX
 XX Novel human diagnostic protein #7336.
 DE
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 KW
 XX Homo sapiens.
 OS
 XX MO200175067-A2.
 PN
 XX

```

PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT.
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS71532.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 37704; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 95 AA;
SQ
XX
XX Query Match 47.6%; Score 39; DB 22; Length 95;
XX Best Local Similarity 77.8%; Pred. No. 27;
XX Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 4 KTRNGFRVP 12
Db 77 KTRNGFRIP 85
XX
XX RESULT 15
XX ABG02544
XX ID ABG02544 standard; Protein; 243 AA.
XX
XX AC ABG02544;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #2535.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX

```

```

PF 30-MAR-2001; 2001MO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT.
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS66731.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 32903; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 243 AA;
SQ
XX
XX Query Match 47.6%; Score 39; DB 22; Length 243;
XX Best Local Similarity 77.8%; Pred. No. 73;
XX Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 4 KTRNGFRVP 12
Db 225 KTRNGFRIP 233
XX
XX Search completed: July 30, 2003, 16:23:30
XX Job time : 33.5238 secs
XX

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 13.5714 Seconds
(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-8
Perfect score: 47
Sequence: 1 APRKNVRW 8

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: SPTEMBL_23:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organellar:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*
16: sp_virus:*
17: sp_bacteriap:*
17: sp_archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	100.0	33	6	Q9TR80
2	41	87.2	266	5	Q9BKM0
3	38	80.9	48	6	Q9BKM0
4	38	80.9	704	6	Q8WMN8
5	38	80.9	1072	5	Q8T104
6	37	78.7	314	11	Q8K2K0
7	37	78.7	555	4	Q96N65
8	37	78.7	555	11	Q8BMA3
9	37	78.7	1126	5	Q9BPR0
10	37	78.7	1712	5	Q8ICM1
11	36	76.6	166	5	Q8BKM0
12	36	76.6	238	5	Q8BKM0
13	36	76.6	1051	12	Q66108
14	35	74.5	345	5	P91109
15	35	74.5	372	2	O05723
16	35	74.5	379	16	Q98CW3

17	35	74.5	379	16	Q8YQ02	Q8YQ02 bruceella me
18	35	74.5	379	16	Q8G161	Q8G161 bruceella su
19	35	74.5	404	16	Q9Y882	Q9Y882 staphylococ
20	35	74.5	404	16	Q8CQ90	Q8CQ90 staphylococ
21	35	74.5	484	17	Q8TN52	Q8TN52 methanosarc
22	35	74.5	496	15	Q99412	Q99412 human immun
23	35	74.5	2513	12	Q87644	Q87644 sindbis vir
24	34	72.3	96	5	Q8IJB8	Q8IJB8 plasmodium
25	34	72.3	118	10	Q9AS17	Q9AS17 oryza sativ
26	34	72.3	122	12	Q64873	Q64873 human adeno
27	34	72.3	256	4	Q9BKM0	Q9BKM0 homo sapien
28	34	72.3	351	12	Q8B8T6	Q8B8T6 human adeno
29	34	72.3	356	16	Q8UY81	Q8UY81 simian adeno
30	34	72.3	380	16	Q9F2R4	Q9F2R4 streptomyce
31	34	72.3	410	17	Q8TZP3	Q8TZP3 pyrococcus
32	34	72.3	445	10	Q8H086	Q8H086 oryza sativ
33	34	72.3	462	15	Q72608	Q72608 human immun
34	34	72.3	491	16	Q9FNB7	Q9FNB7 corynebacte
35	34	72.3	500	15	Q9WQ10	Q9WQ10 human immun
36	34	72.3	518	16	Q8XBZ1	Q8XBZ1 chlorobium
37	34	72.3	1032	13	Q8AXZ4	Q8AXZ4 brachydantio
38	34	72.3	1277	12	Q9JGJ7	Q9JGJ7 sorghum chl
39	33	70.2	85	12	Q98433	Q98433 paramicum
40	33	70.2	164	5	Q17570	Q17570 caenorhabdi
41	33	70.2	183	16	Q9KXJ3	Q9KXJ3 streptomyce
42	33	70.2	185	5	Q94783	Q94783 trypanosoma
43	33	70.2	202	16	Q934V8	Q934V8 salmoneilla
44	33	70.2	227	5	Q18487	Q18487 caenorhabdi
45	33	70.2	244	12	Q41888	Q41888 asparagus v

ALIGNMENTS

RESULT 1

```

ID Q9TR80 PRELIMINARY; PRT; 33 AA.
AC Q9TR80;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE Lactoferrin (Fragment).
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_Taxid=9940;
RN [1]
RP SEQUENCE.
RX MEDLINE=55127729; PubMed=7827104;
RA Qian Z.Y., Jolles P., Migliore-Samour D., Fiat A.M.,
RL Biochim. Biophys. Acta 1243:25-32(1995).
DR HSSP; 077698; 1CE2.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 1.
SQ SEQUENCE 33 AA; 3914 MW; DL904CARL5A73961 CRC64;

```

Query Match 100.0%; Score 47; DB 6; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
DB 1 APRKNVRW 8

RESULT 2

```

ID Q9BKM0 PRELIMINARY; PRT; 266 AA.
AC Q9BKM0;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

```

DE Hypothetical protein Y92H12BL.4.
 GN Y92H12BL.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelioidinae; Caenorhabditis.
 OX NCB1_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA Waterston R.;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Tin-Wollam A., Woldmann P., Courtney L.;
 RT "The sequence of C. elegans cornid Y92H12BL.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RA Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC084268; AAK19005.3; -
 KW Hypothetical protein.
 SQ SEQUENCE 266 AA; 30750 MW; 4CA2328D7899AAB5 CRC64;

Query Match 87.2%; Score 41; DB 5; Length 266;
 Best Local Similarity 75.0%; Pred. No. 2.3;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKXVW 8
 DB 109 APRKXVW 116

RESULT 3
 077558 PRELIMINARY; PRT; 48 AA.
 AC 077558;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Lactoferrin (Fragment).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCB1_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=98332734; PubMed=966128;
 RA Wang S.-R., Lin J., Cheng I.C., Lin T.Y.;
 RT "Characterization and functional analysis of the porcine lactoferrin
 RT gene promoter.";
 RL Gene 215:203-212(1998).
 DR EMBL: AF044256; AAC34369.1; -
 DR InterPro: IPR001156; Transferrin.
 DR Pfam: PF00405; Transferrin; 1.
 FT NON_TER
 SQ SEQUENCE 48 AA; 5569 MW; 28403BE7DE144D78 CRC64;

Query Match 80.9%; Score 38; DB 6; Length 48;
 Best Local Similarity 75.0%; Pred. No. 1.7;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKXVW 8
 DB 20 APRKXVW 27

RESULT 4

OSMN8
 ID OSMN8 PRELIMINARY; PRT; 704 AA.
 AC OSMN8;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Lactoferrin (Fragment).
 GN LTF.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCB1_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Wang S.-R., Lin T.-Y., Weng C.-N.;
 RT "Isolation and expression of porcine milk lactoferrin.";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: TRANSFERRIN IS IRON BINDING PROTEIN WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
 CC - SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC EMBL: L77887; AAL40161.1; -
 DR InterPro: IPR001156; Transferrin.
 DR Pfam: PF00405; transferrin; 2.
 DR PRINTS: PR00422; TRANSFERRIN.
 DR SMART: SM00094; TR_FER; 2.
 DR PROSITE: PS00205; TRANSFERRIN_1; 2.
 DR PROSITE: PS00206; TRANSFERRIN_2; 2.
 DR PROSITE: PS00207; TRANSFERRIN_3; 1.
 KW Glycoprotein; Iron transport; Metal-binding; Transport.
 FT NON_TER
 SQ SEQUENCE 704 AA; 77681 MW; 64E769F7503C32 CRC64;

Query Match 80.9%; Score 38; DB 6; Length 704;
 Best Local Similarity 75.0%; Pred. No. 26;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKXVW 8
 DB 20 APRKXVW 27

RESULT 5
 08T104 PRELIMINARY; PRT; 1072 AA.
 AC 08T104;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Projectin-like protein.
 DE BMPROJECTIN.
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
 OC Bombycidae; Bombyx.
 OX NCB1_TaxID=7091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ps0; TISSUE=Posterior silk gland;
 RA Koike Y., Shimada T., Suzuki M.G., Mita K., Abe H., Maeda S.,
 RA Osoegawa K., Devong P.J.;
 RT "Genomic sequence of 320kb containing a kettin orthologue on the 2
 RT chromosome in Bombyx mori.";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ps0; TISSUE=Posterior silk gland;
 RA Koike Y., Shimada T., Suzuki M.G., Mita K., Abe H., Maeda S.,
 RA Osoegawa K., Devong P.J.;
 RT "Genomic sequence of 320kb containing a kettin orthologue on the 2
 RT chromosome in Bombyx mori.";
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB079864; BAB85195.1; --
 DR EMBL: AB090307; BAC10616.1; --
 DR InterPro: IPR003961; FN_11.
 DR InterPro: IPR007110; Ig_Like.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00041; fn3.2.
 DR Pfam: PF00047; Ig.1.
 DR PROSITE: PSS0835; IG_Like.2.
 SQ SEQUENCE 1072 AA; 122004 MW; C7052490D28B29E CRC64;

Query Match 80.9%; Score 38; DB 5; Length 1072;
 Best Local Similarity 85.7%; Pred. No. 41;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRKXVW 8
 DB 14 PRKXVW 20

RESULT 6

Q8K2KO PRELIMINARY; PRT; 314 AA.

AC Q8K2KO;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Similar to maguin-2, maguin-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC031194; AA031194.1; --
 DR InterPro: IPR001478; PDZ.1.
 DR Pfam: PF00595; PDZ.1.
 DR PROSITE: PSS0106; PDZ.1.
 SQ SEQUENCE 314 AA; 34995 MW; 7A263D3AP0B51F79 CRC64;

Query Match 78.7%; Score 37; DB 11; Length 314;
 Best Local Similarity 75.0%; Pred. No. 19;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRXVW 8
 DB 61 APRXVW 68

RESULT 7

Q96N65 PRELIMINARY; PRT; 555 AA.

AC Q96N65;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical protein FLJ31349.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ieshashi T., Kanehori K., Yosida M., Watanabe S., Iehida S., Ono Y.,
 RA Horita T., Hirata S., Murakawa K., Takiguchi S., Kusano J.,
 RA Watanabe M., Fujimori K., Tanai H., Iehida M., Yamashita H., Chiba Y.,
 RA Sugiyama T., Irie K., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Salto K., Nishikawa T.,
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Watanabe M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.,
 RA "NEDO human cDNA sequencing project.";

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

DR EMBL: AK055911; BAB71044.1; --
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001660; SAM.
 DR Pfam: PF00595; PDZ.1.
 DR Pfam: PF00536; SAM.1.
 DR SMART: SM00228; PDZ.1.
 DR SMART: SM00454; SAM.1.
 DR PROSITE: PSS0106; PDZ.1.
 DR PROSITE: PSS0105; SAM_DOMAIN.1.
 KW Hypothetical protein.
 SQ SEQUENCE 555 AA; 61891 MW; 6EF900E174D91C08 CRC64;

Query Match 78.7%; Score 37; DB 4; Length 555;
 Best Local Similarity 75.0%; Pred. No. 33;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRXVW 8
 DB 302 APRXVW 309

RESULT 8

Q8BMA3 PRELIMINARY; PRT; 555 AA.

AC Q8BMA3;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Weakly similar to membrane-associated guanylate kinase-interacting
 DE protein 2 MAGUIN-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=C57BL/6J; TISSUE=Mullerian duct includes surrounding region;
 RC MEDLINE=22354683; PubMed=12466851;
 RX THE FANTOM Consortium.
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 DR EMBL: AK033015; BAC28127.1; --
 SQ SEQUENCE 555 AA; 61869 MW; 341AP9A4425904F2 CRC64;

Query Match 78.7%; Score 37; DB 11; Length 555;
 Best Local Similarity 75.0%; Pred. No. 33;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRXVW 8
 DB 302 APRXVW 309

RESULT 9

Q9BPRO PRELIMINARY; PRT; 1126 AA.

AC Q9BPRO;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Receptor type guanylyl cyclase (EC 4.6.1.2) (Guanylate
 DE cyclase).
 CN Bn-GCY.
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Periygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditypsia; Bombycoidea;
 OC Bombycidae; Bombyx.
 OX NCBI_TaxID=7091;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Shintarou T., Shigeki S., Taeko S., Yasuhisa E., Takaaki N.;
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21376156; Pubmed=11493433; Endo Y., Nishioaka T.;
 RA Tanoue S., Sumida S., Suetsugu T.,
 RT "Identification of a receptor type guanylyl cyclase in the antennal
 RL lobe and antennal sensory neurons of the silkworm, *Bombyx mori*.";
 CC Insect Biochem. Mol. Biol. 31:971-979(2001).
 CC -1- CATALYTIC ACTIVITY: GTP = 3',5'-CYCLIC GMP + DIPHOSPHATE.
 CC -1- CATALYTIC ACTIVITY: BPLONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL
 CC CYCLASE FAMILY.
 CC EMBL; AB047558; BAB32672.1; -.
 DR HSSP; P26769; 1AB8.
 DR InterPro; IPR001828; ANF_receptor.
 DR InterPro; IPR001054; G_cyclase.
 DR InterPro; IPR001993; Mitoch_carrier.
 DR InterPro; IPR001170; Ntpep_receptor.
 DR InterPro; IPR000719; Prot_kinase.
 DR Pfam; PF01094; ANF_receptor; 1.
 DR Pfam; PF00211; guanylate_cyc; 2.
 DR Pfam; PF00069; pkinase; 1.
 DR PRINTS; PRO00255; NATPEPTDER.
 DR PRINTS; PRO0109; TYRKINASE.
 DR PRODOM; PD000001; Prot_kinase; 1.
 DR SMART; SM00044; CYCC; 1.
 DR PROSITE; PS00452; GUANYLYL_CYCLASES_1; 1.
 DR PROSITE; PS50125; GUANYLYL_CYCLASES_2; 1.
 DR PROSITE; PS00215; MITOCH_CARRIER; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR ATP-binding; Kinase; Lyase; Transferase; Tyrosine-protein kinase;
 KW CGMP synthetase;
 KW SEQUENCE 1126 AA; 126212 MW; F6989C580B0A9488 CRC64;

Query Match 78.7%; Score 37; DB 5; Length 1126;
 Best Local Similarity 62.5%; Pred. No. 69;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKXVRW 8
 DB 1077 APRKXVRW 1084

RESULT 10
 OSICAL PRELIMINARY; PRT; 1712 AA.
 ID OSICAL;
 AC OSICAL;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN MAL6P1.157.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OC NCBI_TaxID=36329;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,
 RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
 RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL844505; CAD50541.1; -.
 KW Hypothetical protein.
 KW SEQUENCE 1712 AA; 198276 MW; B2DB83661670036C CRC64;

Query Match 78.7%; Score 37; DB 5; Length 1712;
 Best Local Similarity 85.7%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRKXVRW 8
 DB 1077 PRKXVRW 1084

DB 256 PRKXVRW 262

RESULT 11
 OSRC9 PRELIMINARY; PRT; 166 AA.
 ID OSRC9;
 AC OSRC9;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Protein involved in RETRIEVAL of ER membrane proteins from the early
 DE golgi compartment.
 GN EC008.0700.
 OS Enecephalitozoon cuniculi.
 OC Eukaryota; Fungi; Microsporidia; Unikaryontidae; Enecephalitozoon.
 OC NCBI_TaxID=6035;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=GB-M1;
 RA GenomeScope;
 RA Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RX MEDLINE=21576510; Pubmed=11719806;
 RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thonart F.,
 RA Prensier F., Barbe V., Peyretallade E., Broctier P., Wincker P.,
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
 RA Weissenbach J., Vivas C.P.;
 RT "Genome sequence and gene compaction of the eukaryote parasite
 RL Enecephalitozoon cuniculi.";
 RT Nature 414:450-453(2001).
 DR EMBL; AL590448; CAD26375.1; -.
 DR InterPro; IPR004932; Ret1.
 DR Pfam; PF03248; Ret1; 1.
 KW SEQUENCE 166 AA; 19690 MW; 51D58C48E03F3A31 CRC64;

Query Match 76.6%; Score 36; DB 5; Length 166;
 Best Local Similarity 75.0%; Pred. No. 15;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKXVRW 8
 DB 15 APRKXVRW 22

RESULT 12
 OSYH1 PRELIMINARY; PRT; 238 AA.
 ID OSYH1;
 AC OSYH1;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE RE62711P (CG32442-PA).
 GN BCDA:RE62711 OR CG32442.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=Berkeley;
 RA Stapleton M., Brokstein P., Hong L., Asgavani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarini H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nuno U., Pacleb U., Paragas V., Park S.,
 RA Patel S., Phouaneavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Ceiniker S.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; Pubmed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanosides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.P.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abail J.F., Agbayan A., An H.U., Andrews-Panico C., Baldwin D.,
 RA Ballew R.M., Baau A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Borzhan M.R., Bouck J., Brockstein P., Broctier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Caley S., Dahlke C., Davenport L.B., Davle P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Poser C., Gabriellian A.E., Gary N.S., Galbart W.M., Glaeser K.,
 RA Glodok A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler P., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svayrskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodard, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 [3]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanosides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnak D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Gary N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jatali M., Kruse D., Li P., Matei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svayrskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [4]
 RP SEQUENCE FROM N.A.
 RA Miara S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochick S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise B., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnick J., Whitfield E.,
 RA Ashburner M., Galbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [5]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [6]
 RP SEQUENCE FROM N.A.
 RA Flybase;

RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY071557; M449179.1; -;
 DR EMBL; AE003594; M441216.1; -;
 DR Flybase; FBgn0047079; BCDNA:R52711.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR Pfam; PF04161; Arvi; 1.
 DR PROSITE; PS00190; CYTOCHROME C; 1.
 SQ SEQUENCE 238 AA; 27748 MW; 4A57CE830C4B641 CRC64;

 Query Match 76.6%; Score 36; DB 5; Length 238;
 Best Local Similarity 75.0%; Pred. No. 22;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 QY 1 APRKVRW 8
 DB 204 ATKRLRW 211

 RESULT 13
 ID Q66108 PRELIMINARY; PRT; 1051 AA.
 AC Q66108;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
 DE Citrus leaf rugose ilarivirus putative viral replicase.
 OS Citrus leaf rugose ilarivirus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;
 OC Ilarivirus.
 OC NCBI_TaxID=37126;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96112216; PubMed=8447536;
 RA Scott S.W., Ge X.;
 RT "The nucleotide sequence of citrus leaf rugose virus RNA 1.";
 RL J. Gen. Virol. 76:3233-3238(1995).
 DR EMBL; U23715; AAC49126.1; -;
 DR InterPro; IPR000606; Viral_helicase.
 DR InterPro; IPR002588; V_methyltransfer.
 DR Pfam; PF01443; Viral_helicase1; 1.
 DR Pfam; PF01660; Vmethyltransfer1.
 SQ SEQUENCE 1051 AA; 118326 MW; A686A74981D9859F CRC64;

 Query Match 76.6%; Score 36; DB 12; Length 1051;
 Best Local Similarity 85.7%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

 QY 2 PRKVRW 8
 DB 231 PRLNVRW 237

 RESULT 14
 ID P91109 PRELIMINARY; PRT; 345 AA.
 AC P91109;
 DT 01-MAY-1997 (TEMBLrel. 03, Created)
 DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE Hypothetical 40.6 kDa protein.
 GN C32B5.16.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pseudocercariae; Caenorhabditidae.
 OC NCBI_TaxID=6239;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Br15c01 N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode *C. elegans*: a platform for
 RT investigating biology. The *C. elegans* Sequencing Consortium.";
 RL Science 282:2012-2016(1998).

Search completed: July 30, 2003, 16:29:30
 Job time : 16.5714 secs

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol NZ;
 RA Schneet P.;
 RT "The sequence of *C. elegans* cosmid C32B5.";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol NZ;
 RA Waterston R.;
 RT "Direct Submission";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U80843; AAB37971.1; -
 DR WormPep; C32B5.16; CE08524.
 KW Hypothetical protein.
 SQ SEQUENCE 345 AA; 40557 MW; 1BC0264AE08F32B4 CRC64;

Query Match 74.5%; Score 35; DB 5; Length 345;
 Best Local Similarity 83.3%; Pred. No. 53;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 RKNVNW 8
 |||:
 Db 100 RKNIRW 105

RESULT 15
 ID 005723 PRELIMINARY; PRT; 372 AA.
 AC 005723;
 DT 01-JUL-1997 (TREMBlrel. 04, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Integrase.
 GN INT.
 OS Bacteroides fragilis.
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Bacteroidaceae; Bacteroides.
 OX NCBI_TaxID=817;
 RN [1]
 RP SEQUENCE OF 1-224 FROM N.A.
 RC TRANSPONSON=Tn4555;
 RX MEDLINE=97252504; PubMed=9098073;
 RA Tribble G.D., Parker A.C., Smith C.J.;
 RT "The Bacteroides mobilizable transposon Tn4555 integrates by a site-specific recombination mechanism similar to that of the gram-positive bacterial element Tn916."
 RT J. Bacteriol. 179:2731-2739 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TRANSPONSON=Tn4555;
 RX MEDLINE=99343846; PubMed=10413660;
 RA Tribble G.D., Parker A.C., Smith C.J.;
 RT "Genetic structure and transcriptional analysis of a mobilizable, antibiotic resistance transposon from Bacteroides.";
 RT Plasmid 42:1-12(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TRANSPONSON=Tn4555;
 RA Tribble G.D., Parker A.C., Smith C.J.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U75371; AAB53787.2; -
 DR InterPro; IPR02104; Phage_integrase.
 DR Pfam; PF00589; Phage_integrase; 1.
 SQ SEQUENCE 372 AA; 43472 MW; F2F08AFAF31ACD04D CRC64;

Query Match 74.5%; Score 35; DB 2; Length 372;
 Best Local Similarity 62.5%; Pred. No. 57;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 APRKNVNW 8
 |||:
 Db 104 ADRKNIKW 111

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:10:10 ; Search time 2.61905 Seconds
(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-8
Perfect score: 47
Sequence: 1 APRKRVW 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	100.0	708	1	TRFL_BOVIN
2	47	100.0	708	1	TRFL_BUBBU
3	47	100.0	708	1	TRFL_CAPHU
4	42	89.4	695	1	TRFL_HORSE
5	38	80.9	704	1	TRFL_PIG
6	35	74.5	2512	1	POIN_SINDV
7	34	72.3	347	1	VCOM_ADEL2
8	34	72.3	358	1	VCOM_ADEB4
9	34	72.3	368	1	VCOM_ADE05
10	34	72.3	369	1	VCOM_ADE02
11	34	72.3	439	1	RT30_HUMAN
12	34	70.2	166	1	LITR_HUMAN
13	33	70.2	166	1	LITR_HUMAN
14	33	70.2	194	1	R15E_ACFPU
15	33	70.2	326	1	BTUC_SALTI
16	33	70.2	326	1	BTUC_SALTY
17	33	70.2	382	1	YH6E_YEAST
18	33	70.2	404	1	OAL_HUMAN
19	33	70.2	694	1	LCF3_YEAST
20	33	70.2	1451	1	BM30_ASAIH
21	33	70.2	3726	1	TRX_DROME
22	33	68.1	96	1	YS98_MYCLE
23	32	68.1	442	1	RT30_MOUSE
24	32	68.1	538	1	C1WA_PAT
25	32	68.1	694	1	LCF4_YEAST
26	32	68.1	741	1	CH12_YEAST
27	32	68.1	2236	1	PYR1_DROME
28	31	67.0	510	1	YOR3_CAEEL
29	31	66.0	118	1	VE4_HPV13
30	31	66.0	124	1	Y212_STRCO
31	31	66.0	189	1	MAP_BACSU
32	31	66.0	198	1	ALCD_HUMAN
33	31	66.0	198	1	ALCD_MOUSE

34	31	66.0	326	1	BTUC_ECO57	Q8x417 escherichia
35	31	66.0	326	1	BTUC_ECOLI	P06609 escherichia
36	31	66.0	350	1	C1WF_HUMAN	Q9h427 homo sapien
37	31	66.0	357	1	VNS2_BTV10	P23065 bluetongue
38	31	66.0	357	1	VNS2_BTV1X	P10350 bluetongue
39	31	66.0	408	1	PILC_VIBCH	Q9x499 vibrio chol
40	31	66.0	447	1	YRFL_ECOLI	P39319 escherichia
41	31	66.0	468	1	SG3_HUMAN	Q8wx32 homo sapien
42	31	66.0	471	1	SG3_MOUSE	P47867 mus musculu
43	31	66.0	471	1	SG3_RAT	P47868 ratu8 norv
44	31	66.0	507	1	N04C_MESVI	Q8um86 mesocytiga
45	31	66.0	515	1	Y314_BUCAP	Q8x9m0 buchneera ap

ALIGNMENTS

RESULT 1

ID	TRFL_BOVIN	STANDARD	PRT	708 AA.
AC	P24627; Q29629; Q9WZV3;			
DT	01-MAR-1992 (Rel. 21, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Lactoferrin precursor (Lactoferrin) [Contains: Lactoferritin B			
DE	(LFCIN B)]			
GN	LTF			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_Taxid=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Submaxillary gland;			
RX	MEDLINE=9160550; PubMed=2001696;			
RX	Pierce A., Colavizza D., Benalissa M., Maes P., Tartar A.,			
RA	Montreuil J., Spik G.;			
RT	"Molecular cloning and sequence analysis of bovine lactoferrin";			
RL	Eur. J. Biochem. 1961;177:184(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92028986; PubMed=1718281;			
RA	Goodman R.E., Schanbacher F.L.;			
RT	"Bovine lactoferrin mRNA: sequence, analysis, and expression in the			
RL	mammary gland.";			
RN	Biochem. Biophys. Res. Commun. 180:75-84(1991).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Lung;			
RA	Tsang T.C., Burns D.K., Wang F., Pan Y.C.E., Schmidt A.M., Stern D.;			
RT	"Cloning of a 80-kD advanced glycosylation end product (AGE) binding			
RL	protein from bovine lung.";			
RN	FASEB J. 6:233-233(1991).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Blood, and Mammary gland;			
RX	MEDLINE=9426164; PubMed=8206385;			
RA	Seyfert H.-W., Tuckoritz A., Interthal H., Koczan D., Hobom G.;			
RT	"Structure of the bovine lactoferrin-encoding gene and its promoter.";			
RL	Gene 143:265-269(1994).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RA	Nakamura I., Shimazaki K., Yagi Y., Watanabe A.;			
RN	Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.			
RN	[6]			
RP	SEQUENCE OF 20-59.			
RX	MEDLINE=90031466; PubMed=2805645;			
RA	Rejman J.J., Hegarty H.M., Hurley W.L.;			
RT	"Purification and characterization of bovine lactoferrin from			
RL	secretions of the involuting mammary gland: identification of			
RT	multiple molecular weight forms.";			
RL	Comp. Biochem. Physiol. 93B:929-934(1989).			

Query Match 100.0%; Score 47; DB 1; Length 708;
 Best Local Similarity 100.0%; Pred. No. 0.059;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 APRKNVW 8
 |||||
 20 APRKNVW 27

RESULT 2

TRFL_BUBBU STANDARD; PRT; 708 AA.
 AC 07698;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Lactoferrin precursor (lactoferrin).
 CN LTP.
 OS Bubalus bubalis (Domestic water buffalo).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bubalus.
 OX NCBI_TaxID=89462;
 [1]
 RP SEQUENCE FROM N.A.
 RA Paramesivam M., Thattaiyath B.D., Kumar A., Srinivasan A.,
 RA Singh T.P.;
 RT "cDNA sequence of Buffalo lactoferrin.";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 [2]
 RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE=2003130; Pubmed=10531476;
 RA Karthikeyan S., Paramesivam M., Yadav S., Srinivasan A., Singh T.P.;
 RT "Structure of buffalo lactoferrin at 2.5-A resolution using crystals
 grown at 303 K shows different orientations of the N and C lobes.";
 RL Acta Crystallogr. D 55:1805-1813(1999).
 CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 OF AN ANION, USUALLY BICARBONATE.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement. (See <http://www.isb-sdb.cn/announce/>
 or send an email to license@isb-sdb.cn).
 CC -----
 CC EMBL; AJ005203; CAA06441.1; -.
 DR PDB; 1CE2; 19-MAR-99.
 DR PDB; 1BIY; 13-JAN-99.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 2.
 KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 KW Signal; 3D-structure.
 FT SIGNAL 1 19
 FT CHAIN 20 708 LACTOTRANSFERRIN.
 FT DISULFID 28 64
 FT DISULFID 38 55
 FT DISULFID 134 217
 FT DISULFID 176 192
 FT DISULFID 179 202
 FT DISULFID 189 200

FT DISULFID 250 264
 FT DISULFID 367 399
 FT DISULFID 377 390
 FT DISULFID 424 703
 FT DISULFID 444 666
 FT DISULFID 476 551
 FT DISULFID 500 694
 FT DISULFID 510 524
 FT DISULFID 521 534
 FT DISULFID 592 606
 FT DISULFID 644 649
 FT METAL 79
 FT METAL 111
 FT METAL 111
 FT METAL 211
 FT METAL 272
 FT METAL 414
 FT METAL 452
 FT METAL 452
 FT METAL 545
 FT METAL 545
 FT BINDING 140
 FT BINDING 140
 FT BINDING 482
 FT CARBOHYD 252
 FT CARBOHYD 300
 FT CARBOHYD 495
 FT CARBOHYD 564
 FT TURN 21
 FT TURN 22
 FT STRAND 25
 FT STRAND 32
 FT HELIX 46
 FT STRAND 47
 FT STRAND 53
 FT HELIX 61
 FT TURN 70
 FT STRAND 75
 FT HELIX 80
 FT TURN 89
 FT STRAND 93
 FT STRAND 102
 FT STRAND 107
 FT STRAND 110
 FT TURN 119
 FT HELIX 125
 FT TURN 129
 FT STRAND 132
 FT STRAND 135
 FT TURN 138
 FT TURN 141
 FT HELIX 145
 FT TURN 151
 FT HELIX 152
 FT TURN 159
 FT HELIX 164
 FT TURN 170
 FT STRAND 174
 FT TURN 176
 FT TURN 178
 FT TURN 182
 FT HELIX 186
 FT TURN 189
 FT TURN 190
 FT HELIX 196
 FT TURN 199
 FT TURN 200
 FT TURN 204
 FT TURN 205
 FT TURN 208
 FT HELIX 211
 FT TURN 219
 FT TURN 220
 FT STRAND 221
 FT STRAND 225
 FT TURN 230
 FT HELIX 232
 FT HELIX 240
 FT TURN 244
 FT STRAND 245
 FT TURN 249
 FT TURN 251
 FT STRAND 252
 FT STRAND 255
 FT HELIX 257
 FT TURN 260
 FT HELIX 261
 FT TURN 263
 FT STRAND 267
 FT STRAND 270
 FT STRAND 273
 FT STRAND 277

IRON 1.
 IRON 1.
 IRON 1.
 IRON 1.
 IRON 2.
 IRON 2.
 IRON 2.
 ANION (POTENTIAL).
 ANION (POTENTIAL).
 N-LINKED (GLYCAC. . .) (POTENTIAL).
 N-LINKED (GLYCAC. . .) (POTENTIAL).
 N-LINKED (GLYCAC. . .) (POTENTIAL).
 N-LINKED (GLYCAC. . .) (POTENTIAL).

```

FT HELIX 283 297
PT TURN 289 300
FT TURN 307 308
PT TURN 312 313
FT TURN 321 322
PT STRAND 325 328
FT TURN 331 332
PT HELIX 335 335
FT HELIX 354 351
PT STRAND 364 369
PT HELIX 371 384
PT TURN 385 385
PT STRAND 388 393
FT HELIX 396 404
PT TURN 405 406
PT STRAND 410 413
PT HELIX 415 422
PT TURN 423 425
PT STRAND 426 434
PT TURN 438 441
PT TURN 444 446
PT STRAND 452 459
PT TURN 460 461
PT HELIX 463 464
PT HELIX 471 472
PT STRAND 474 477
PT TURN 480 481
FT TURN 483 486
FT HELIX 487 497
PT TURN 503 505
PT STRAND 508 510
PT TURN 512 513
PT TURN 516 517
PT TURN 519 520
PT TURN 522 523
PT TURN 528 529
PT TURN 533 534
PT TURN 538 539
PT TURN 541 542
PT HELIX 544 553
PT TURN 554 555
PT STRAND 559 563
FT HELIX 564 569
PT TURN 570 570
FT HELIX 578 581
PT TURN 582 582
PT TURN 585 586
PT STRAND 588 591
PT TURN 593 594
PT STRAND 597 599
PT TURN 600 601
PT TURN 603 605
PT STRAND 609 611
PT STRAND 615 619
PT HELIX 620 637
PT TURN 639 640
PT TURN 642 647
PT TURN 650 651
PT TURN 654 655
PT TURN 662 663
PT STRAND 664 669
PT HELIX 676 680
PT HELIX 682 692
PT TURN 693 694
PT HELIX 698 705

```

Query Match 100.0%; Score 47; DB 1; Length 708;
 Best Local Similarity 100.0%; Pred. No. 0.059; 0;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 APRKVRW 8
 |||||

```

DB 20 APRKVRW 27
RESULT 3
TRFL_CAPHI STANDARD; PRT; 708 AA.
AC Q29477; Q29479;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactoferrin precursor (Lactoferrin).
GN LTF.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Lee T., Yu S., Kim S., Lee K., Yu D.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA MEDLINE=94380047; PubMed=6093048;
RL Provost P., Nocart M., Guerin G., Martin P.;
RT "Characterization of the goat lactoferrin cDNA. Assignment of the
RT relevant locus to bovine U12 syntenic group.";
RL Biochem. Biophys. Res. Commun. 203:1324-1332(1994).
CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U53857; AAA97958.1; -
DR EMBL; X78902; CA55517.1; -
DR HSPG; O77698; ICE2.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin.2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SM00094; TR_FER; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
KW Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 708 LACTOTRANSFERRIN.
FT REPEAT 20 363 1.
FT REPEAT 364 708 2.
FT DISULFID 28 64 BY SIMILARITY.
FT DISULFID 38 55 BY SIMILARITY.
FT DISULFID 134 217 BY SIMILARITY.
FT DISULFID 176 192 BY SIMILARITY.
FT DISULFID 189 200 BY SIMILARITY.
FT DISULFID 250 264 BY SIMILARITY.
FT DISULFID 367 399 BY SIMILARITY.
FT DISULFID 377 390 BY SIMILARITY.
FT DISULFID 424 703 BY SIMILARITY.
FT DISULFID 444 666 BY SIMILARITY.
FT DISULFID 476 551 BY SIMILARITY.

```

```

FT DISULFID 500 694 BY SIMILARITY.
FT DISULFID 510 524 BY SIMILARITY.
FT DISULFID 521 534 BY SIMILARITY.
FT DISULFID 592 606 BY SIMILARITY.
FT DISULFID 644 649 BY SIMILARITY.
FT METAL 79 79 IRON 1 (BY SIMILARITY).
FT METAL 111 111 IRON 1 (BY SIMILARITY).
FT METAL 211 211 IRON 1 (BY SIMILARITY).
FT METAL 272 272 IRON 1 (BY SIMILARITY).
FT METAL 414 414 IRON 2 (BY SIMILARITY).
FT METAL 452 452 IRON 2 (BY SIMILARITY).
FT METAL 545 545 IRON 2 (BY SIMILARITY).
FT METAL 614 614 IRON 2 (BY SIMILARITY).
FT BINDING 140 140 ANION (BY SIMILARITY).
FT BINDING 482 482 ANION (BY SIMILARITY).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 56 56 I -> V (IN REF. 2).
FT CONFLICT 88 88 Q -> R (IN REF. 2).
FT CONFLICT 124 124 Q -> R (IN REF. 2).
FT CONFLICT 154 154 F -> P (IN REF. 2).
FT CONFLICT 304 304 S -> R (IN REF. 2).
FT CONFLICT 414 414 D -> G (IN REF. 2).
SQ SEQUENCE 708 AA; 77358 MW; F2EDA3C83539960D CRC64;

```

```

Query Match 100.0%; Score 47; DB 1; Length 708;
Best Local Similarity 100.0%; Pred. No. 0.059;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 APRKVRM 8
DB 20 APRKVRM 27

```

```

RESULT 4
TRFL_HORSE STANDARD; PRT; 695 AA.

```

```

AC 07811;
DT 15-JUN-1999 (Rel. 38, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lactoferrin precursor (Lactoferrin) (Fragment).
GN LTF.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RA Paramasivam M., Srinivasan A., Yadav M.P., Singh T.P.;
RT "cDNA sequence of mare lactoferrin."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RC TISSUE=Milk;
RX MEDLINE=99296631; PubMed=10366507;
RA Sharma A.K., Paramasivam M., Srinivasan A., Yadav M.P., Singh T.P.;
RT "Three-dimensional structure of mare deferitic lactoferrin at 2.6-A
resolution."
J. Mol. Biol. 289:303-317(1999).
CC - FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION. USUALLY BICARBONATE.
CC - SUBUNIT: Monomer.
CC - SUBCELLULAR LOCATION: Secreted.
CC - DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC - SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; AJ010930; CA09407.1; -
CC PDB; 1B1X; 02-DEC-98.
CC PDB; 1B7U; 02-FEB-99.
CC PDB; 1B7Z; 02-FEB-99.
CC PDB; 1F9B; 10-FEB-01.
CC PDB; 1F9B; 13-FEB-02.
CC PDB; 1QJW; 14-JAN-00.
CC InterPro; IPR001156; Transferrin.
CC Pfam; PF00405; transferrin; 2.
CC PRINTS; PR00422; TRANSFERRIN.
CC SMART; SM00094; TR_FER; 2.
CC PROSITE; PS00205; TRANSFERRIN_1; 2.
CC PROSITE; PS00206; TRANSFERRIN_2; 2.
CC PROSITE; PS00207; TRANSFERRIN_3; 1.
CC Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
CC Signal; 3D-structure.
CC KX Signal; 3D-structure.
CC KW NON TER 1 6
CC FT SIGNAL 1 6
CC FT CHAIN 7 695
CC FT REPEAT 7 350
CC FT REPEAT 351 695
CC FT DISULFID 15 51
CC FT DISULFID 25 42
CC FT DISULFID 121 204
CC FT DISULFID 163 179
CC FT DISULFID 166 189
CC FT DISULFID 176 187
CC FT DISULFID 237 251
CC FT DISULFID 354 386
CC FT DISULFID 364 377
CC FT DISULFID 411 690
CC FT DISULFID 431 653
CC FT DISULFID 463 538
CC FT DISULFID 487 681
CC FT DISULFID 497 511
CC FT DISULFID 508 521
CC FT DISULFID 579 593
CC FT DISULFID 631 636
CC FT METAL 66 66
CC FT METAL 98 98
CC FT METAL 198 198
CC FT METAL 259 259
CC FT METAL 401 401
CC FT METAL 433 433
CC FT METAL 532 532
CC FT METAL 601 601
CC FT BINDING 127 127
CC FT BINDING 469 469
CC FT CARBOHYD 143 143
CC FT CARBOHYD 287 287
CC FT CARBOHYD 482 482
CC FT STRAND 12 16
CC FT STRAND 19 34
CC FT TURN 35 36
CC FT STRAND 40 44
CC FT HELIX 48 56
CC FT TURN 57 58
CC FT STRAND 62 65
CC FT HELIX 67 74
CC FT TURN 76 78
CC FT STRAND 80 88
CC FT STRAND 95 95
CC FT STRAND 97 105
CC FT TURN 106 107
CC FT HELIX 112 114
CC FT TURN 116 117
CC FT STRAND 119 122

```

```

FT      TURN      125      126
PT      TURN      128      131
PT      HELIX      132      137
PT      HELIX      139      142
PT      TURN      147      148
PT      HELIX      151      158
PT      STRAND      161      163
PT      TURN      165      166
PT      TURN      169      171
PT      HELIX      173      175
PT      TURN      176      178
PT      TURN      183      187
PT      TURN      191      192
PT      TURN      194      195
PT      HELIX      197      206
PT      TURN      207      208
PT      STRAND      212      216
PT      HELIX      219      223
PT      HELIX      227      230
PT      TURN      231      232
PT      STRAND      233      237
PT      TURN      238      240
PT      STRAND      241      244
PT      HELIX      245      247
PT      TURN      248      250
PT      STRAND      254      257
PT      STRAND      260      264
PT      HELIX      270      284
PT      TURN      286      287
PT      TURN      294      295
PT      TURN      308      309
PT      STRAND      312      315
PT      HELIX      322      326
PT      HELIX      328      336
PT      TURN      337      338
PT      HELIX      341      349
PT      STRAND      351      356
PT      HELIX      358      370
PT      TURN      371      372
PT      STRAND      375      380
PT      HELIX      383      392
PT      TURN      393      393
PT      STRAND      397      400
PT      HELIX      402      410
PT      TURN      411      412
PT      STRAND      421      421
PT      HELIX      429      433
PT      STRAND      439      446
PT      HELIX      454      456
PT      TURN      458      459
PT      STRAND      462      464
PT      TURN      467      468
PT      TURN      470      473
PT      HELIX      474      484
PT      HELIX      489      491
PT      TURN      492      492
PT      STRAND      496      497
PT      TURN      499      500
PT      TURN      503      504
PT      TURN      506      507
PT      STRAND      508      508
PT      TURN      515      516
PT      TURN      520      521
PT      TURN      525      526
PT      TURN      528      529
PT      HELIX      531      540
PT      TURN      541      542
PT      STRAND      546      550
PT      HELIX      551      555
PT      TURN      556      557
PT      TURN      559      560
PT      HELIX      565      568
PT      TURN      569      569

```

```

FT      TURN      572      573
PT      STRAND      575      578
PT      TURN      580      581
PT      STRAND      584      586
PT      HELIX      587      592
PT      STRAND      596      598
PT      STRAND      602      605
PT      TURN      607      609
PT      HELIX      610      624
PT      TURN      626      627
PT      TURN      629      630
PT      HELIX      631      634
PT      TURN      637      638
PT      TURN      641      642
PT      TURN      649      650
PT      STRAND      651      655
PT      TURN      658      659
PT      HELIX      663      667
PT      HELIX      669      679
PT      TURN      680      681

Query Match      89.4%; Score 42; DB 1; Length 695;
Best Local Similarity 87.5%; Pred. No. 0.62;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 APRKRVW 8
DB      7 APRKRVW 14

RESULT 5
TRFL_PIG      STANDARD; PRT; 704 AA.
ID      ID      PIG      STANDARD; PRT; 704 AA.
AC      P14632; 029557;
DT      01-APR-1980 (Rel. 14, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Lactoferrin precursor (Lactoferrin).
GN      LTF.
OS      Sus scrofa (Pig).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX      NCBI_TaxID=9623;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92379101; PubMed=1511016;
RA      Lyndon J.P., O'Malley B.R., Saucedo O., Lee T., Headon D.R.,
RA      Connolly O.M.;
RT      "Nucleotide and primary amino acid sequence of porcine lactoferrin.";
RL      Biochim. Biophys. Acta 1132:97-99(1992).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92367939; PubMed=1503259;
RA      Alexander L.J., Levine W.B., Teng C.T., Beattie C.W.;
RT      "Cloning and sequencing of the porcine lactoferrin cDNA.";
RN      [3]
RP      SEQUENCE OF 20-49.
RX      MEDLINE=90105538; PubMed=2605266;
RA      Hutchens T.W., Magnuson J.S., Hip T.-T.;
RT      "Rapid purification of porcine colostrum whey lactoferrin by affinity
RT      chromatography on single-stranded DNA-agarose. Characterization,
RT      amino acid composition and N-terminal amino acid sequence.";
RL      Biochim. Biophys. Acta 999:323-329(1989).
CC      -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC      CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC      OF AN ANION, USUALLY BICARBONATE.
CC      -1- SUBUNIT: Monomer.
CC      -1- SUBCELLULAR LOCATION: Secreted.
CC      -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC      -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.ldb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC -----
 CC EMBL; M92089; AAA31102.1; -;
 CC EMBL; M81327; AAA31059.1; -;
 CC PIR; A45543; A45543.
 CC HSSP; 077698; 1C82.
 CC InterPro: IPR001155; Transferrin.
 CC Pfam; PF004405; Transferrin; 2.
 CC PRINTS; PR00422; TRANSFERRIN.
 CC SMART; SM00094; TR_FER; 2.
 CC PROSITE; PS00205; TRANSFERRIN_1; 2.
 CC PROSITE; PS00206; TRANSFERRIN_2; 2.
 CC PROSITE; PS00207; TRANSFERRIN_3; 1.
 CC Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 KW Signal.
 FT 1 19
 FT CHAIN 20 704
 FT REPEAT 20 359
 FT REPEAT 360 704
 FT DISULFID 28 62
 FT DISULFID 38 53
 FT DISULFID 130 213
 FT DISULFID 172 188
 FT DISULFID 185 196
 FT DISULFID 246 260
 FT DISULFID 363 395
 FT DISULFID 373 386
 FT DISULFID 420 699
 FT DISULFID 472 547
 FT DISULFID 496 690
 FT DISULFID 506 520
 FT DISULFID 517 530
 FT DISULFID 588 602
 FT DISULFID 640 645
 FT METAL 77 77
 FT METAL 107 107
 FT METAL 207 268
 FT METAL 268 268
 FT METAL 410 410
 FT METAL 448 448
 FT METAL 541 541
 FT METAL 610 610
 FT BINDING 136 136
 FT BINDING 478 478
 FT CARBOHYD 385 385
 FT CARBOHYD 491 491
 FT CONFLICT 12 12
 FT CONFLICT 46 48
 FT CONFLICT 51 51
 FT CONFLICT 85 85
 FT CONFLICT 121 121
 FT CONFLICT 132 132
 FT CONFLICT 284 284
 FT CONFLICT 573 573
 FT CONFLICT 590 590
 FT CONFLICT 625 625
 FT CONFLICT 662 662
 FT CONFLICT 704 704
 FT CONFLICT 7625 7625
 FT SEQUENCE 704 AA; 7625 MW; 93261EFD608AD358 CRC64;

Query Match 80.9%; Score 38; DB 1; Length 704;
 Best Local Similarity 75.0%; Pred. No. 4.2;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 APRKVRW 8
 DB 20 APRKVRW 27

RESULT 6
 ID POLN_SINDV STANDARD; PRT; 2512 AA.
 AC P03317;
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nonstructural polypeptide (P270) [Contains: Nonstructural protein
 DE NSP3; Nonstructural NSP2; Nonstructural NSP4].
 OS Sindbis virus (strain HRSF).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 NCBI_TaxID=11034;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8414839; PubMed=6322438;
 RA Straus E.G., Rice C.M., Straus J.H.;
 RT "Complete nucleotide sequence of the genomic RNA of Sindbis virus.";
 RN Virology 133:92-110(1984).
 RN [2]
 RP SEQUENCE OF 1-54 FROM N.A.
 RX MEDLINE=83268700; PubMed=6308269;
 RA Ou J.H., Straus E.G., Straus J.H.;
 RT "The 5'-terminal sequences of the genomic RNAs of several
 RT alphaviruses.";
 RN J. Mol. Biol. 168:1-15(1983).
 RN [3]
 RP SEQUENCE OF 1429-2512 FROM N.A.
 RX MEDLINE=83299955; PubMed=6577423;
 RA Straus E.G., Rice C.M., Straus J.H.;
 RT "Sequence coding for the alphavirus nonstructural proteins is
 RT interrupted by an open termination codon.";
 RN Proc. Natl. Acad. Sci. U.S.A. 80:5271-5275(1983).
 RN [4]
 RP SEQUENCE OF 2431-2512 FROM N.A.
 RX MEDLINE=83039346; PubMed=6291034;
 RA Ou J.H., Rice C.M., Dalgarno L., Straus E.G., Straus J.H.;
 RT "Sequence studies of several alphavirus genomic RNAs in the region
 RT containing the start of the subgenomic RNA.";
 RN Proc. Natl. Acad. Sci. U.S.A. 79:5235-5239(1982).
 CC -1- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.
 CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UGA OCCURS
 CC BETWEEN THE CODONS FOR 1896-TYR AND 1897-LEU.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.ldb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC -----
 CC EMBL; J02363; AAA96975.1; ALT_SEQ.
 CC PIR; A03917; MNWVS.
 CC MEROPS; C09_001; -;
 DR InterPro: IPR002588; ALP.
 DR InterPro: IPR002620; Peptidase_C9.
 DR InterPro: IPR001788; RNA_dep_RNAPol2.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_Psivir.
 DR InterPro: IPR000606; Viral_helicase1.
 DR Pfam; PF01661; Alpp; 1.
 DR Pfam; PF01707; Peptidase_C9; 1.
 DR Pfam; PF00978; RNA_dep_RNAPol2; 1.
 DR Pfam; PF01443; Viral_helicase1; 1.
 DR SMART; SM00506; Alpp; 1.
 KW Polypeptide; Nonstructural protein; RNA-binding; Helicase.
 FT CHAIN 1 540
 FT CHAIN 541 1347
 FT CHAIN 1348 1896
 FT CHAIN 1897 2512
 FT NONSTRUCTURAL PROTEIN NSP1.
 FT NONSTRUCTURAL PROTEIN NSP2.
 FT NONSTRUCTURAL PROTEIN NSP3.
 FT NONSTRUCTURAL PROTEIN NSP4.

SQ SEQUENCE 2512 AA; 279546 MW; F3656FCBBBA95726 CRC64;
 Query Match 74.5%; Score 35; DB 1; Length 2512;
 Best Local Similarity 62.5%; Pred. No. 64;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 APRKXVW 8
 Db 1195 APRKRIEM 1202
 RESULT 7
 VCOM_ADE12 STANDARD; PRT; 347 AA.
 AC P36717;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Minor core protein (Protein V).
 GN PV.
 OS Human adenovirus type 12.
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 OX NCBI_TaxID=28282;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94076430; PubMed=8254750;
 RA Sprangel J., Schmitz B., Heuss-Weitzel D., Zock C., Doerfler W.;
 RT "Nucleotide sequence of human adenovirus type 12 DNA: comparative
 RT functional analysis.";
 RL J. Virol. 68:379-389(1994).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; X73487; CAAS188.1; -
 CC DR PIR; S33939; S33939.
 CC DR InterPro; IPR005608; Adeno_PV.
 CC DR Pfam; PF03910; Adeno_PV; 1.
 CC KW Core protein; Late protein.
 CC SQ SEQUENCE 347 AA; 39369 MW; 3C338F62AAS3027 CRC64;
 Query Match 72.3%; Score 34; DB 1; Length 347;
 Best Local Similarity 62.5%; Pred. No. 13;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 APRKXVW 8
 Db 61 APRRVQW 68
 RESULT 8
 VCOM_ADE40 STANDARD; PRT; 358 AA.
 AC P48753;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Minor core protein (Protein V).
 GN PV.
 OS Human adenovirus type 40.
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 OX NCBI_TaxID=28284;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=Dugan;
 RX MEDLINE=94087748; PubMed=8263936;
 RA Davison A.J., Telford E.A., Watson M.S., McBride K., Maurer V.;
 RT "The DNA sequence of adenovirus type 40.";

RL J. Mol. Biol. 234:1308-1316(1993).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; L19443; AAC13964.1; -
 CC DR InterPro; IPR005608; Adeno_PV.
 CC DR Pfam; PF03910; Adeno_PV; 1.
 CC KW Core protein; Late protein.
 CC SQ SEQUENCE 358 AA; 40707 MW; B7D0D3BAF794C968 CRC64;
 Query Match 72.3%; Score 34; DB 1; Length 358;
 Best Local Similarity 62.5%; Pred. No. 14;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 APRKXVW 8
 Db 63 APRRVQW 70
 RESULT 9
 VCOM_ADE05 STANDARD; PRT; 368 AA.
 ID VCOM_ADE05
 AC P24938;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Minor core protein (Protein V).
 GN PV.
 OS Human adenovirus type 5.
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 OX NCBI_TaxID=28285;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92087470; PubMed=1727603;
 RA Chroboczek J., Bieber F., Jacrot B.;
 RT "The sequence of the genome of adenovirus type 5 and its comparison
 RT with the genome of adenovirus type 2.";
 RL Virology 186:280-285(1992).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; M73260; AAA96409.1; -
 CC DR PIR; C39449; FOADM5.
 CC DR InterPro; IPR005608; Adeno_PV.
 CC DR Pfam; PF03910; Adeno_PV; 1.
 CC KW Core protein; Late protein.
 CC SQ SEQUENCE 368 AA; 41446 MW; 722E6C6D22C692M4 CRC64;
 Query Match 72.3%; Score 34; DB 1; Length 368;
 Best Local Similarity 62.5%; Pred. No. 14;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 APRKXVW 8
 Db 59 APRRVQW 66
 RESULT 10
 VCOM_ADE02 STANDARD; PRT; 369 AA.
 ID VCOM_ADE02
 AC P03257;

```

DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Minor core protein (Protein V).
GN PV.
OS Human adenovirus type 2.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=10515;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=55054835; PubMed=6094534;
RA Alsterlund P., Akusjall G., Lager M., Yeh-Kai L., Pettersson U.,
RL J. Biol. Chem. 259:113980-113985 (1984).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J01917; AAA92213.1; -
DR PIR; A03837; FOADM2.
DR InterPro: IPR005608; Adeno_PV.
DR Pfam; PF03910; Adeno_PV; 1.
RM Core protein; Late protein.
SQ SEQUENCE 369 AA; 41721 MW; 577C9E645B5B57DE CRC64;

Query Match 72.3%; Score 34; DB 1; Length 369;
Best Local Similarity 62.5%; Pred. No. 14;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 APRRQVM 8
DB 60 APRRQVM 67

RESULT 11
RT30_HUMAN STANDARD; PRT: 439 AA.
ID Q9NP52; Q9NP52; Q9NP52; Q9NP52; Q9NP52;
AC Q9NP52; Q9NP52; Q9NP52; Q9NP52; Q9NP52;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Mitochondrial 28S ribosomal protein S30 (S30mt) (MRP-S30) (Programmed
DE cell death protein 9) (BM047).
GN MRP530 OR PDCD9.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
RA MEDLINE=20108793; PubMed=10640817;
RA Carim L., Sunoy L., Nadal M., Battivili X., Becarcelier M.;
RA "Cloning, expression, and mapping of PDCD9, the human homolog of
RA Gallus gallus pro-apoptotic protein p52";
RL Cytogenet. Cell Genet. 87:85-88 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Carim L., Battivili X., Becarcelier M., Sunoy L.;
RA Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Giesel S.,
RA Ansoerge W., Boecker M., Blocher H., Baerisachs S., Blum H.,
RA Lauber J., Duesterhoelt A., Beyer A., Koehler K., Strack N.,
RA Mewes H.-W., Oltmannseder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;

```

```

RT "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs."
RL Genome Res. 11:1422-1435 (2001).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Ito T., Ota T., Nishikawa T., Hayaishi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hiro Y., Saito K.,
RA Yamamoto Y., Nakamatsu A., Nakamura Y., Kojima S., Nagatani K.,
RA Masuno Y., Ono T., Okumura K., Yoshikawa T., Aoyama S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Nishimura K.;
RT "NDO human cDNA sequencing project";
RN Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=2338257; PubMed=12477932;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shewen C.W., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carrinci P., Prange C.J.,
RA Raha S.S., Lonnellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunnarone P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywicki M.T., Skalska U., Smalins D.E.,
RA Schermer A., Schein J.E., Jones S.J.W., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [6]
RP SEQUENCE OF 89-439 FROM N.A.
RC TISSUE=Bone marrow;
RA Zhao M., Gu J., Li N., Peng Y., Han Z., Chen Z.;
RT "A novel gene expressed in human bone marrow";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
RP IDENTIFICATION.
RX MEDLINE=21276436; PubMed=11279123;
RA Koc E.C., Burkhart W., Blackburn K., Moseley A., Sprengli L.L.;
RT "The small subunit of the mammalian mitochondrial ribosome:
RT identification of the full complement of ribosomal proteins present."
RL J. Biol. Chem. 276:11963-11974 (2001).
RN [8]
RP -1-SUBUNIT: Component of the mitochondrial ribosome small subunit
RN (28S) which comprises a 12S rRNA and about 30 distinct proteins.
CC -1-SUBUNIT: Component of the mitochondrial ribosome.
CC -1-TISSUE SPECIFICITY: Heart, skeletal muscle, kidney and liver.
CC -1-EXPRESSION: Expression in placenta and peripheral blood leukocytes.
CC -1-CAUTION: Ref. 6 sequence differs from that shown due to a number of
CC sequencing errors that include frameshifts.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF146192; AAF65227.1; -
CC EMBL; AL355715; CAB90810.1; -
CC EMBL; AL355716; CAB90811.1; -
CC EMBL; AL35706; CAB66641.1; -
CC EMBL; AK074777; BAC11202.1; -
CC EMBL; BC007735; AA07735.1; -
CC EMBL; AF217523; AAF67634.1; ALT_FRAME.

```

DR Genew: HGNC:8769; MRPS30.
 DR GO: GO:0006915; P:apoptosis; TAS.
 KM Ribosomal protein; Mitochondrion.
 FT COMPACT 33 S -> C (IN REF. 4 AND 5).
 FT COMPACT 82 I -> V (IN REF. 3).
 SO SEQUENCE 439 AA; 50348 MW; AD0564743B03290C CIRC64;
 Query Match 72.3%; Score 34; DB 1; Length 439;
 Best local similarity 71.4%; Pred. No. 17;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 PRKXRM 8
 DB 388 PRKXRM 394
 RESULT 12
 LITR_HUMAN
 ID LITR_HUMAN STANDARD; PRT; 166 AA.
 AC P05451; P11379;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Lithostathine 1 alpha precursor (Pancreatic stone protein) (PSP)
 DE (Pancreatic thread protein) (PTP) (Islet of Langerhans regenerating
 DE protein) (REG) (Regenerating protein I alpha) (Islet cells
 DE regeneration factor) (ICRF).
 GN REGIA OR REG OR PSPS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_Taxid:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88115343; PubMed=296300;
 RA Terazone K., Yamamoto H., Takasawa S., Shiga K., Yonemura Y.,
 RA Tochino Y., Okamoto H.;
 RT "A novel gene activated in regenerating islets.";
 RL J. Biol. Chem. 263:2111-2114(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90237042; PubMed=2332435;
 RA Watanabe T., Yonekura H., Terazone K., Yamamoto H., Okamoto H.;
 RT "Complete nucleotide sequence of human reg gene and its expression in
 RT normal and tumoral tissues. The reg protein, pancreatic stone
 RT protein, and pancreatic thread protein are one and the same product
 RT of the gene.";
 RL J. Biol. Chem. 265:7432-7439(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX TISUE=PANCREAS;
 RC MEDLINE=89292148; PubMed=2525567;
 RA Girotti D., Bernard J.-P., Konquiter S., Iovanna J., Sables H.,
 RA Dagon J.-C.;
 RT "Secretory pancreatic stone protein messenger RNA. Nucleotide
 RT sequence and expression in chronic calcifying pancreatitis.";
 RL J. Clin. Invest. 84:100-106(1989).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX BOONVARTISAWAT W., TANDHANAND-BANCHUIN N., VAMSAENG S.,
 RA YENCHITSOMANUS P.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RX TISUE=PANCREAS;
 RC MEDLINE=22388257; PubMed=12477932;
 RA STRAUBERG R.L., FEINGOLD E.A., GROUSE L.H., DERGE J.G.,
 RA KLAUSNER R.D., COLLINS F.S., WAGNER L., SHERMAN C.M., SCHULER G.D.,
 RA ALTSCHUL S.F., ZEEBERG B., BUETOW K.H., SCHAEFER C.F., BHAT N.K.,
 RA HOPKINS R.F., JORDAN H., MOORE T., MAX S.I., WANG J., HSEIH F.,
 RA DIACHENKO L., MARSHALL K., FARMER A.A., RUBIN G.M., HONG L.,
 RA STAPLETON M., SOARES M.B., BONALDO M.F., CASAVANT T.L., SCHEETZ T.E.,
 RA BROWNSTEIN M.J., UEDIN T.B., TOCHIYUKI S., CARINCI P., PRANGE C.,

RA RANA S.S., LOQUELLANO N.A., PETERE G.J., ADPRAMSON R.D., MULLABY S.J.,
 RA BOSAK S.A., MCBRAN P.J., MCKERNAN K.J., MALEK J.A., GUNNARNE P.H.,
 RA RICHARDS S., WOLEY K.C., HALE S., GARCIA A.M., GAY L.J., HULIK S.W.,
 RA VILLALON D.K., MUZNY D.M., SODERGREN E.U., LU X., GIBBS R.A.,
 RA FAHEY J., HELTON E., KETEMAN M., MADAN A., RODRIGUES S., SANCHEZ A.,
 RA WHITING M., MADAN A., YOUNG A.C., SHEVCHENKO Y., BOUFFARD G.G.,
 RA BLAKELEY R.W., TOUCHMAN J.W., GREEN E.D., DICKSON M.C.,
 RA RODRIGUEZ A.C., GRIMWOOD J., SCHMUTZ J., MYERS R.M.,
 RA BUTTERFIELD Y.S.N., KRZYWINKI M.I., SKALSKA U., SMALLUE D.E.,
 RA SCHNERCH A., SCHEIN J.E., JONES S.J.M., MARRA M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [6]
 RP SEQUENCE OF 34-166.
 RX MEDLINE=88029417; PubMed=3665916;
 RA de Caro A.M., Bonicel J.J., Rouini P., de Caro J.D., Sables H.,
 RA ROVERE M.;
 RT "Complete amino acid sequence of an immunoreactive form of human
 RT pancreatic stone protein isolated from pancreatic juice.";
 RL Eur. J. Biochem. 168:201-207(1987).
 RN [7]
 RP SEQUENCE OF 34-98.
 RX MEDLINE=87099950; PubMed=3541906;
 RA Montalto G., Bonicel J.J., Multigner L., ROVERE M., Sables H.,
 RA de Caro A.M.;
 RT "Partial amino acid sequence of human pancreatic stone protein, a
 RT novel pancreatic secretory protein.";
 RL Biochem. J. 238:227-232(1986).
 RN [8]
 RP SEQUENCE OF 34-78.
 RX MEDLINE=8606356; PubMed=3908481;
 RA Gross J., Carlson R.I., Brauer A.W., Margolies M.N., MARSHAW A.L.,
 RA WANDS J.R.;
 RT "Isolation, characterization, and distribution of an unusual
 RT pancreatic human secretory protein.";
 RL J. Clin. Invest. 76:2115-2125(1985).
 RN [9]
 RP SEQUENCE OF 23-47.
 RX MEDLINE=89150292; PubMed=2493268;
 RA de Caro A.M., Adrich Z., Fournet B., Capon C., Bonicel J.J.,
 RA de Caro J.D., ROVERE M.;
 RT "N-terminal sequence extension in the glycosylated forms of human
 RT pancreatic stone protein. The 5-oxoproline N-terminal chain is O-
 RT glycosylated on the 5th amino acid residue.";
 RL Biochim. Biophys. Acta 994:281-284(1989).
 RN [10]
 RP SEQUENCE OF 33-58.
 RX MEDLINE=87219142; PubMed=3108036;
 RA Rouini P., Bonicel J., ROVERE M., de Caro A.;
 RT "Cleavage of the Arg-Ile bond in the native polypeptide chain of
 RT human pancreatic stone protein.";
 RL FEBS Lett. 216:195-199(1987).
 RN [11]
 RP IDENTITY OF REG WITH PSP.
 RX MEDLINE=89350859; PubMed=2764894;
 RA Stewart T.A.;
 RT "The human reg gene encodes pancreatic stone protein.";
 RL Biochem. J. 260:622-623(1989).
 RN [12]
 RP DISULFIDE BONDS.
 RX MEDLINE=91032149; PubMed=2226837;
 RA Itoh T., Tsuzuki H., Kato T., Teraoka H., Matsumoto K., Yoshida N.,
 RA Terazone K., Watanabe T., Yonekura H., Yamamoto H., Okamoto H.;
 RT "Isolation and characterization of human reg protein produced in
 RT Saccharomyces cerevisiae.";
 RL FEBS Lett. 272:85-88(1990).
 RN [13]
 RP ALZHEIMER'S DISEASE AND DEVELOPMENTAL EXPRESSION REGULATION.
 RX MEDLINE=90368981; PubMed=2394826;
 RA de la Monte S.M., Ozturk M., Wands J.R.;
 RT "Enhanced expression of an exocrine pancreatic protein in Alzheimer's
 RT disease and the developing human brain.";

RL J. Clin. Invest. 86:1004-1013(1990).
 RN [14]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
 RX MEDLINE=96256285; PubMed=8654365;
 RA Bertrand J.A., Pignol D., Bernard J.-P., Verdier J.-M., Dagorn J.-C.,
 RA Fontecilla-Camps J.C.;
 RT "Crystal structure of human lithostathine, the pancreatic inhibitor
 RT of stone formation.";
 RL EMBL J. 15:2678-2684(1996).
 RN [15]
 RP X-RAY CRYSTALLOGRAPHY (1.30 ANGSTROMS) OF 23-166.
 RX MEDLINE=20092874; PubMed=10625646;
 RA Gerdaud V., Pignol D., Loret E., Bertrand J.A., Berland Y.,
 RA Fontecilla-Camps J.C., Canselier J.P., Gabas N., Verdier J.M.;
 RT "Mechanism of calcite crystal growth inhibition by the N-terminal
 RT undecapeptide of lithostathine.";
 RL J. Biol. Chem. 275:1057-1064(2000).
 RN [16]
 RP STRUCTURE BY NMR OF 34-164.
 RX MEDLINE=97120677; PubMed=8661348;
 RA Patard L., Stoven V., Gharib B., Bontems F., Lallemand J.-Y.,
 RA de Reggi M.;
 RT "What function for human lithostathine? structural investigations by
 RT three-dimensional structure modeling and high-resolution NMR
 RT spectroscopy.";
 RL Protein Eng. 9:949-957(1996).
 CC -1- FUNCTION: MIGHT ACT AS AN INHIBITOR OF SPONTANEOUS CALCIUM
 CC CARBONATE PRECIPITATION. MAY BE ASSOCIATED WITH NEURONAL
 CC SPROUTING IN BRAIN, AND WITH BRAIN AND PANCREAS REGENERATION.
 CC -1- TISSUE SPECIFICITY: IN PANCREATIC ACINAR CELLS AND, IN LOWER
 CC LEVELS, IN BRAIN.
 CC -1- DEVELOPMENTAL STAGE: HIGH EXPRESSION LEVELS IN FETAL AND INFANT
 CC BRAINS; MUCH LOWER IN ADULT BRAINS.
 CC -1- DISEASE: ALZHEIMER'S DISEASE AND DOWN'S SYNDROME PATIENTS SHOW
 CC ENHANCED EXPRESSION OF PSP-RELATED TRANSCRIPTS AND INTRANEUROAL
 CC ACCUMULATION OF PSP-LIKE PROTEINS IN THEIR BRAINS.
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M27190; AAA60546.1; -;
 DR EMBL; M27189; AAA60545.1; -;
 DR EMBL; M18963; AAA36558.1; -;
 DR EMBL; J05412; AAA36559.1; -;
 DR EMBL; AF172331; AAD51330.1; -;
 DR EMBL; BC005350; AAH05350.1; -;
 DR PIR; A35197; RGH01A.
 DR PIR; A45751; A45751.
 DR PDB; 1LIT; 11-JAN-97.
 DR PDB; 1QDD; 24-JAN-01.
 DR Genew; HGNC:9951; REG1A.
 DR MIM; 167770; -;
 DR MIM; 167800; -;
 DR GO; GO:0008284; P:positive regulation of cell proliferation; TAS.
 DR InterPro; IPR001353; AntiFreezeZell.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR003990; Pancreatic_ac.
 DR Pfam; PF00059; lectin_c; 1.
 DR PRINTS; PR01504; PNCGRATTISAP.
 DR PRINTS; PR00356; ANTIFREEZEZEL.
 DR SMART; SMO0034; CLECT.1.
 DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
 DR PROSITE; PS00615; C_TYPE_LLECTIN_2; 1.
 DR Glycoprotein; Signal; Alzheimer's disease; Down's syndrome; Lectin;
 KW 3D-structure; Pyridone carboxylic acid.
 FT SIGNAL 1 22

Query Match 70.2%; Score 33; DB 1; Length 166;
 Best Local Similarity 71.4%; Pred. No. 10;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Oy 2 PRKRRW 8
 Db 105 PRKRRW 111

RESULT 13

ID	LIB	HUMAN	STANDARD	PRT	166 AA.
AC	P48304				
DT	01-FEB-1996	Rel. 33, Created			
DT	01-FEB-1996	Rel. 33, Last sequence update			
DT	28-FEB-2003	Rel. 41, Last annotation update			
DE	Lithostathine 1 beta precursor (Regenerating protein I beta).				
GN	REG1B OR REG1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Pancreas.				
RX	MEDLINE=94153997; PubMed=8110835;				
RA	Motilzumi S., Watanabe T., Uno M., Nakagawara K.I., Suzuki Y.,				
RA	Miyashita H., Yonekura H., Okamoto H.;				
RT	"Isolation, structural determination and expression of a novel reg				
RT	gene, human reg1 beta.";				
RL	Biochim. Biophys. Acta 1217:199-202(1994).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=93351647; PubMed=8348956;				
RA	Bartoli C., Gharib B., Giorgi D., Sansonetti A., Dagorn J.-C.,				
RA	Berge-Lefranc U.;				
RT	"A gene homologous to the reg gene is expressed in the human				
RT	pancreas.";				
RL	FEBS Lett. 327:289-293(1993).				
RN	[3]				
RP	CARBOHYDRATE-LINKAGE SITE.				
RX	MEDLINE=95331286; PubMed=7607222;				
RA	De Reggi M., Capon C., Gharib B., Wieruszski J.M., Michel R.,				
RA	Fournet B.;				
RT	"The glycan moiety of human pancreatic lithostathine. Structure				
RT	characterization and possible pathophysiological implications.";				
RL	Eur. J. Biochem. 230:503-510(1995).				
CC	-1- FUNCTION: MIGHT ACT AS AN INHIBITOR OF SPONTANEOUS CALCIUM				
CC	CARBONATE PRECIPITATION. MAY BE ASSOCIATED WITH NEURONAL				
CC	SPROUTING IN BRAIN, AND WITH BRAIN AND PANCREAS REGENERATION.				
CC	-1- PWM: ALL O-LINKED GLYCANS CONSIST OF GAL-GALNAc-GAL-GALNAc				
CC	TETRASACCHARIDE CORE AND GET ELONGATED (MICROHETEROGENITY).				
CC	-1- SIMILARITY: Contains 1 C-type lectin family domain.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). CC -----				
DR	EMBL; D17291; BAA04124.1; -;				
DR	EMBL; D16816; BAA04091.1; -;				
DR	EMBL; L08080; AAA18204.1; -;				
DR	PIR; S34591; RGH01B.				
DR	HSSP; P05451; 1LIT.				
DR	Genew; HGNC:9952; REG1B.				
DR	MIM; 167771; -;				
DR	GO; GO:0008283; P:cell proliferation; TAS.				
DR	InterPro; IPR001304; Lectin_C.				
DR	InterPro; IPR003990; Pancreatic_ac.				
DR	Pfam; PF00059; lectin_c; 1.				

```

DR PRINTS: PR01504; PNCREATTISAP.
DR SMART: SMO0034; CLECT. 1.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS00041; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Signal; Lectin; Pyroglutamate carboxylic acid.
FT SIGNAL 1 22
FT CHAIN 1 166
FT DOMAIN 34 164
FT MOD_RES 23 23
FT CARBOHYD 27 27
FT DISULFID 36 47
FT DISULFID 64 162
FT DISULFID 137 154
SQ SEQUENCE 166 AA; 18665 MW; D1DC20E11AE5DDE8 CRC64;

Query Match 70.2%; Score 33; DB 1; Length 166;
Best Local Similarity 71.4%; Pred. No. 10;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRKXRV 8
DB 105 PKKRRW 111

RESULT 14
RISE_ARCFU STANDARD; PRT; 194 AA.
AC 027965;
AD 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L15e.
GN RPL15E OR AP2319.
OS Archaeoglobus fulgidus.
OC Archaeae; Buryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_Taxid=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleen H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gaitan M., Hickey E.K., Peterson J.D.,
RA Rischdschman D.L., Kervatue A.R., Graham D.E., Kyriades N.C.,
RA Kitzman E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.R., Badger J.H., Glodok A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uitterback T.,
RA Corton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujit C., Garland S.A.,
RA Mason T.M., Olsen G.J., Frazer C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- SIMILARITY: BELONGS TO THE L15E FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000944; AAB88937.1; -.
CC PIR; G69539; G69539.
CC TIGR; AF2319; -.
CC HAMAP; MF_00256; -; 1.
CC InterPro; IPR000439; Ribosomal_L15e.
CC Pfam; PF00827; Ribosomal_L15e; 1.
CC PROSITE; PS01194; RIBOSOMAL_L15e; 1.

```

```

KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 194 AA; 23167 MW; 4B607B60F132BD23 CRC64;

Query Match 70.2%; Score 33; DB 1; Length 194;
Best Local Similarity 57.1%; Pred. No. 12;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 PRKXRV 8
DB 93 PKKRRW 99

RESULT 15
RUC_SALTI STANDARD; PRT; 326 AA.
ID BUC_SALTI
AC 082615;
AD 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Vitamin B12 transport system permease protein bnc.
GN BUC OR STY1770 OR T1221.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_Taxid=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=1167608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,
RA Baker S., Basham A., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar P.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Lio S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
CC -1- FUNCTION. Part of the binding-protein-dependent transport system
CC for vitamin B12. Involved in the translocation of the substrate
CC across the membrane (By similarity).
CC -1- SUBCELLULAR LOCATION. Integral membrane protein. Inner membrane
CC (Probable).
CC -1- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM PERMEASE FAMILY. FECCD SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL627271; CAD02012.1; -.
CC EMBL; AE016838; AAO68676.1; -.
CC HAMAP; MF_01004; -; 1.
CC InterPro; IPR000515; BPD_transp.
CC InterPro; IPR000522; FECCD.
CC Pfam; PF01032; FECCD; 1.
CC PRODOM; PD001557; FECCD; 1.
CC PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; FALSE_NEG.

```

KW Transport; Cobalt transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 15 35 POTENTIAL.
FT TRANSMEM 61 81 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 113 133 POTENTIAL.
FT TRANSMEM 146 166 POTENTIAL.
FT TRANSMEM 184 204 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 274 294 POTENTIAL.
FT TRANSMEM 302 322 POTENTIAL.
SQ SEQUENCE 326 AA; 34864 MW; 3835896331225CB CRC64;

Query Match 70.2%; Score 33; DB 1; Length 326;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Oy 3 RKNVRW 8
Db 10 RKNVRW 15

Search completed: July 30, 2003, 16:24:33
Job time : 3.61905 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:55 ; Search time 5.09524 Seconds
(without alignments)
150.994 Million cell updates/sec

Title: US-09-787-070-8

Perfect score: 47

Sequence: 1 APRKNVRW 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.76:*
2: p1r1:*
3: p1r2:*
4: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	100.0	33	2 S52107	lactoferrin - sheep
2	47	100.0	708	2 TFBOL	lactotransferrin p
3	47	100.0	708	2 JC2323	lactoferrin - goat
4	38	80.9	703	2 A45543	lactoferrin precu
5	35	74.5	345	2 T25580	hypothetical prote
6	35	74.5	379	2 AD3390	phm protein [impo
7	35	74.5	404	2 B89819	pyrimidine nucleos
8	35	74.5	2512	1 MNWVS	nonstructural poly
9	34	72.3	347	2 S33939	minor core protein
10	34	72.3	368	1 FOADM2	minor core protein
11	34	72.3	369	1 FOADM2	minor core protein
12	33	70.2	85	2 T17882	hypothetical prote
13	33	70.2	164	2 T30991	hypothetical prote
14	33	70.2	166	1 RGHU1A	regenerating islet
15	33	70.2	166	1 RGHU1B	regenerating islet
16	33	70.2	166	1 A45751	pancreatic stone p
17	33	70.2	194	2 G69539	ribosomal protein
18	33	70.2	227	2 S72574	hypothetical prote
19	33	70.2	253	2 T30381	hypothetical prote
20	33	70.2	326	2 AC0705	vitamin B12 transp
21	33	70.2	466	2 T30040	hypothetical prote
22	33	70.2	468	2 S46791	hypothetical prote
23	33	70.2	489	2 T13026	hypothetical prote
24	33	70.2	694	2 B54901	long-chain-fatty-a
25	33	70.2	776	2 T29064	hyaluronate lyase
26	33	70.2	1451	2 S65571	patern formation
27	33	70.2	1795	2 T30372	avirulence protein
28	33	70.2	3759	2 A35085	trithorax protein
29	32	68.1	56	2 F82830	hypothetical prote

30	32	68.1	96	2 A87110	conserved hypothet
31	32	68.1	153	2 S77791	hypothetical prote
32	32	68.1	245	2 T26868	hypothetical prote
33	32	68.1	245	2 G70985	hypothetical prote
34	32	68.1	276	2 T50896	hypothetical membr
35	32	68.1	330	2 A87194	polyprenyl diphosp
36	32	68.1	366	2 H86425	hypothetical prote
37	32	68.1	367	2 F69816	reticuline oxidase
38	32	68.1	379	2 E95946	probable enzyme pr
39	32	68.1	404	2 A63467	hypothetical prote
40	32	68.1	476	2 D71264	conserved hypothet
41	32	68.1	508	2 C82138	conserved hypothet
42	32	68.1	661	2 C83843	hypothetical prote
43	32	68.1	694	2 S56060	long-chain-fatty-a
44	32	68.1	741	2 S50340	CHL2 protein - Ye
45	32	68.1	771	2 T34376	hypothetical prote

ALIGNMENTS

RESULT 1
S52107
lactoferrin - sheep (fragment)
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C>Date: 14-Jul-1995 #sequence_revision 05-Dec-1997 #text_change 05-Dec-1997
C:Accession: S52107
R:Qian, Z.Y.; Jolles, P.; Migliore-Samouri, D.; Fiat, A.M.
Biochim. Biophys. Acta 1243, 25-32, 1995
A>Title: Isolation and characterization of sheep lactoferrin, an inhibitor of platelet a
A:Reference number: S52107; MUID:95127729; PMID:7827104
A:Accession: S52107
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-33 <Q1a>
C:Superfamily: transferrin; transferrin repeat homology
C:Keywords: duplication

Query Match 100.0%; Score 47; DB 2; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
DB 1 APRKNVRW 8

RESULT 2
TFBOL
lactotransferrin precursor - bovine
N:Alternate names: lactoferrin
C:Species: Bos primigenius taurus (cattle)
C>Date: 31-Mar-1992 #sequence_revision 21-Nov-1997 #text_change 11-May-2000
C:Accession: I45919; S14674; S14110; S18517; J070595; S130518; S13881; P10148; S21
R:Tsang, T.C.; Burns, D.K.; Wang, F.; Pan, Y.
FASEB J. 6, 233, 1991
A>Title: Cloning of a 80-kD advanced glycosylation end product (AGE) binding protein fro
A:Reference number: I45919
A:Accession: I45919
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-708 <TSA>
A:Cross-references: GB:I08604; NID:g163269; PIDN:AAA30609.1; PID:g163270
R:Pierce, A.
submitted to the EMBL Data Library, November 1990
A:Reference number: S14674
A:Accession: S14674
A:Molecule type: mRNA
A:Residues: 1-144/V/V, 146-163, /P/P, 166-339, /A/, 341-438, /Y/, 440-513, /R/, 515-708 <PI>
A:Cross-references: EMBL:X57084; NID:G505; PIDN:CAA40366.1; PID:G506
R:Pierce, A.; Colavizza, D.; Benalissa, M.; Maes, P.; Tartar, A.; Montreuil, J.; Spik, G
Eur. J. Biochem. 196, 177-184, 1991
A>Title: Molecular cloning and sequence analysis of bovine lactotransferrin.

A:Reference number: S14110; MUID:91160550; PMID:2001696
 A:Accession: S14110
 A:Molecule type: mRNA
 A:Residues: 3-144, 'V', 146-339, 'A', 341-438, 'Y', 440-513, 'R', 515-708 <PI2>
 A:Cross-references: EMBL:X57084
 A:Accession: S18517
 A:Molecule type: protein
 A:Residues: 20-35;82-114;148-163, 'P', 166-178, 'V', 'P', 183-190;205-212;230-239;304-339;59
 R:Goodman, R.E.; Schanbacher, P.L.
 Biochem. Biophys. Res. Commun. 180, 75-84, 1991
 A:Title: Bovine lactoferrin mRNA: sequence, analysis, and expression in the mammary gland
 A:Reference number: J05955; MUID:92028986; PMID:1718281
 A:Accession: J05955
 A:Molecule type: mRNA
 A:Residues: 1-65, 'PG', 68-296, 'S', 298-339, 'A', 341-708 <GOO>
 A:Cross-references: GB:M63502
 A:More: the authors translated the codon CCG for residue 66 as Arg and TCT for residue 2
 R:Mead, P.E.; Tweedie, J.W.
 Nucleic Acids Res. 18, 7167, 1990
 A:Title: cDNA and protein sequence of bovine lactoferrin.
 A:Reference number: S13097; MUID:91088328; PMID:2263492
 A:Accession: S13097
 A:Molecule type: mRNA
 A:Residues: 28-33, 'DS', 36-38, 'P', 40-708 <MEA>
 A:Cross-references: EMBL:X54801
 A:Accession: S18518
 A:Molecule type: protein
 A:Residues: 20-47;59-66;132-139;256-277;278,305-332;343-351;361-363;586,587-589;598-619
 R:Mead, P.E.
 submitted to the EMBL Data Library, October 1990
 A:Reference number: S13881
 A:Accession: S13881
 A:Molecule type: mRNA
 A:Residues: 28-38, 'P', 40-86, 'C', 88-708 <ME3>
 A:Cross-references: EMBL:X54801
 R:Reifman, J.J.; Hegarty, H.M.; Hurley, W.L.
 Comp. Biochem. Physiol. B 93, 929-934, 1989
 A:Title: Purification and characterization of bovine lactoferrin from secretions of the
 A:Reference number: P10148; MUID:90031466; PMID:2805645
 A:Accession: P10148
 A:Molecule type: protein
 A:Residues: 20-27, 'X', 29-37, 'X', 39-54, 'X', 56-59 <BEJ>
 R:Bellamy, W.; Takase, M.; Yamuchi, K.; Makabayashi, H.; Kawase, K.; Tomita, M.
 Biochim. Biophys. Acta 1121, 130-136, 1992
 A:Title: Identification of the bactericidal domain of lactoferrin.
 A:Reference number: S21756; MUID:92287941; PMID:1599334
 A:Accession: S21756
 A:Molecule type: protein
 A:Residues: 36-60 <BEJ>
 R:Shimazaki, K.; Tanaka, T.; Kon, H.; Oota, K.; Kawaguchi, A.; Maki, Y.; Sato, T.
 J. Dairy Sci. 76, 946-955, 1993
 A:Title: Separation and characterization of the C-terminal half molecule of bovine lacto
 A:Reference number: A56659; MUID:93253156; PMID:8486845
 A:Accession: A56659
 A:Molecule type: protein
 A:Residues: 20-25;302-308;359-366, 'X', 368-376, 'X', 378 <SHI>
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keywords: duplication; glycoprotein; iron; iron binding; metalloprotein; milk
 F:1-19/Domain: signal sequence #status predicted <Sig>
 F:120-708/Product: lactoferrin repeat homology <TRH>
 F:120-355/Domain: transferrin repeat homology <TRH>
 F:36-60/Region: antimicrobial
 F:359-696/Domain: transferrin repeat homology <TRH>
 F:328-64;134-217;176-192;179-200;189-202;250-264;367-399;377-390;424-703;444-666;476-551,
 F:38-55/Diulfide bonds: #status predicted
 F:79,111,211,272/Binding sites: iron (Asp, Tyr, Tyr, His) #status experimental
 F:140/Binding site: carbonate (Arg) #status experimental
 F:252,300,387,495,564/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:164,455,545,614/Binding site: iron (Asp, Tyr, Tyr, His) #status experimental
 F:482/Binding site: carbonate (Arg) #status experimental
 Query Match 100.0%; Score 47; DB 1; Length 708;
 Beef Local Similarity 100.0%; Pred. No. 0.25;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 APRNRVW 8
 |||||
 Db 20 APRNRVW 27
 RESULT 3
 J02323
 lactoferrin - goat
 C:Species: Capra aegagrus hircus (domestic goat)
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 07-May-1999
 C:Accession: J02323
 C:Keywords: P.; Nocart, M.; Guerin, G.; Martin, P.
 R:Le Provost, P.; Nocart, M.; Guerin, G.; Martin, P.
 Biochem. Biophys. Res. Commun. 203, 1324-1332, 1994
 A:Title: Characterization of the goat lactoferrin cDNA: assignment of the relevant locus
 A:Reference number: J02323; MUID:94380047; PMID:8092048
 A:Accession: J02323
 A:Molecule type: mRNA
 A:Residues: 1-708 <LEP>
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keywords: duplication; glycoprotein
 F:359-696/Domain: transferrin repeat homology <TRH>
 F:252,300,387,495,564/Binding site: carbohydrate (Asn) (covalent) #status predicted
 Query Match 100.0%; Score 47; DB 2; Length 708;
 Beef Local Similarity 100.0%; Pred. No. 0.25; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 APRNRVW 8
 |||||
 Db 20 APRNRVW 27
 RESULT 4
 A45543
 lactoferrin precursor - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 04-Mar-2000
 C:Accession: A45543; S24173
 R:Alexander, L.J.; Levine, W.B.; Teng, C.T.; Beattie, C.W.
 Anim. Genet. 23, 251-256, 1992
 A:Title: Cloning and sequencing of the porcine lactoferrin cDNA.
 A:Reference number: A45543; MUID:92367939; PMID:1501259
 A:Accession: A45543
 A:Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-703 <ALE>
 A:Experimental source: mammary gland
 A:Note: Sequence extracted from NCBI backbone (NCBI:U11151, NCBI:P11153)
 R:Lydon, J.P.; O'Malley, B.R.; Sencido, O.; Lee, T.; Headon, D.R.; Connely, O.M.
 Biochim. Biophys. Acta 1132, 97-99, 1992
 A:Title: Nucleotide and primary amino acid sequence of porcine lactoferrin.
 A:Reference number: S24173; MUID:92379101; PMID:1511016
 A:Accession: S24173
 A:Molecule type: mRNA
 A:Residues: 1-11, 'W', 13-50, 'I', 52-84, 'G', 86-120, 'L', 121-130, 'I', 132-282, 'S', 284-571, 'Q',
 A:Cross-references: EMBL:M92089; NID:g164613; PIDD:AAA1102.1; PID:g164614
 A:Experimental source: mammary gland
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keywords: duplication; glycoprotein; iron; iron binding; metalloprotein; milk
 F:1-19/Domain: signal sequence #status predicted <Sig>
 F:120-703/Product: lactoferrin repeat homology <TRH>
 F:120-350/Domain: transferrin repeat homology <TRH>
 F:36-48/Region: antimicrobial
 F:354-691/Domain: transferrin repeat homology <TRH>
 F:28-62;38-53;129-212;171-187;184-195;245-259;367-394;372-385;419-698;439-661;471-546;49
 F:177,107,206,267/Binding sites: iron (Asp, Tyr, Tyr, His) #status predicted
 F:135/Binding site: carbonate (Arg) #status predicted
 F:409,447,540,609/Binding site: iron (Asp, Tyr, Tyr, His) #status predicted
 F:477/Binding site: carbonate (Arg) #status predicted
 F:490/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 80.9%; Score 38; DB 2; Length 703;
 Best Local Similarity 75.0%; Pred. No. 15;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKXVM 8
 |||:|
 20 APRKXVM 27

RESULT 5

T25580
 hypothetical protein C32B5.16 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T25580

R:Scheet, P.
 submitted to the EMBL Data Library, December 1996

A:Description: The sequence of *C. elegans* cosmid C32B5.

A:Reference number: Z20054

A:Accession: T25580
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-345 <SCG>

A:Cross-references: EMBL:U80843; PIDN:AAB37971.1; GSPDB:GN00020; CESP:C32B5.16

A:Experimental source: strain Bristol N2; clone C32B5

C:Genetics:

A:Gene: CESP:C32B5.16

A:Map position: 2

A:Introns: 24/2; 304/2

C:Superfamily: *Caenorhabditis elegans* hypothetical protein F42G2.3

Query Match 74.5%; Score 35; DB 2; Length 345;
 Best Local Similarity 83.3%; Pred. No. 29;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 RKNTRM 8
 |||:|
 100 RKNTRM 105

RESULT 6

AD3390
 phnM protein [imported] - *Brucella melitensis* (strain 16M)

C:Species: *Brucella melitensis*

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

C:Accession: AD3390

R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, R.; DelVecchio, V.G.; Golestan, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Teles, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*

A:Reference number: AD3392; PMID:11756688

A:Accession: AD3390

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-379 <KTR>

A:Cross-references: GB:AB008917; PIDN:AAL52287.1; PID:G17983077; GSPDB:GN00130

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BME11106

A:Map position: 1

Query Match 74.5%; Score 35; DB 2; Length 379;
 Best Local Similarity 75.0%; Pred. No. 32;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 APRKXVM 8
 |||:|
 66 APRKXVM 73

RESULT 7

B89819

pyrimidine nucleoside transport protein nupc [imported] - *Staphylococcus aureus* (strain

C:Species: *Staphylococcus aureus*

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: B89819

R:Kutoda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc, ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.

A:Reference number: A89758; MUID:2131952; PMID:11418146

A:Accession: B89819

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-404 <KUR>

A:Cross-references: GB:BA000018; PID:G13700411; PIDN:BA841709.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: nupc

C:Superfamily: pyrimidine nucleoside transport protein nupc

Query Match 74.5%; Score 35; DB 2; Length 404;
 Best Local Similarity 83.3%; Pred. No. 34;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 RKNTRM 8
 |||:|
 24 RKNTRM 29

RESULT 8

MNWS

nonstructural polyprotein - *Sindbis virus*

N:Contains: nonstructural NS2; nonstructural protein NS1; nonstructural protein NS3; non

C:Species: *Sindbis virus*

C:Date: 19-Feb-1984 #sequence_revision 03-Aug-1984 #text_change 04-Oct-1996

C:Accession: A03917

R:Straus, E.G.; Rice, C.M.; Straus, J.H.

Virology 133, 92-110, 1984

A:Title: Complete nucleotide sequence of the genomic RNA of *Sindbis virus*.

A:Reference number: A94331; MUID:84148439; PMID:6322438

A:Accession: A03917

A:Molecule type: genomic RNA

A:Residues: 1-2512 <STR>

A:Experimental source: strain HRSP

A>Note: readthrough of the terminator UGA between codons UAC for 1896-Tyr and CUA for 18

C:Superfamily: Semliki Forest virus nonstructural protein

C:Keywords: polyprotein

F:1-540/Product: nonstructural protein NS1 #status predicted <NS1>

F:541-1347/Product: nonstructural protein NS2 #status predicted <NS2>

F:1348-1896/Product: nonstructural protein NS3 #status predicted <NS3>

F:1897-2512/Product: nonstructural protein NS4 #status predicted <NS4>

Query Match 74.5%; Score 35; DB 1; Length 2512;
 Best Local Similarity 62.5%; Pred. No. 2e+02;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 APRKXVM 8
 |||:|
 1195 APRKXVM 1202

RESULT 9

S33939

minor core protein - human adenovirus 12

C:Species: *Human adenovirus 12*

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999

C:Accession: S33939

R:Spiegel, J.
 submitted to the EMBL Data Library, June 1993

A:Reference number: S33939

A:Accession: S33939

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-347 <SPR>

A:Cross-references: EMBL:X73487; NID:G3133361; PIDN:CAA51886.1; PID:G313373
C:Superfamily: adenovirus minor core protein pV

Query Match 72.3%; Score 34; DB 2; Length 347;
Best Local Similarity 62.5%; Pred. No. 47;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVW 8
DB 61 APRRVQM 68

RESULT 10
FOADM5

minor core protein pV - human adenovirus 5
C:Species: Mastadenovirus h5 (human adenovirus 5)

A:Note: host Homo sapiens (man)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jul-1999
C:Accession: G39449
R:Chroboczek, J.; Biebert, F.; Jacrot, B.
Virology 186, 280-285, 1992

A:Title: The sequence of the genome of adenovirus type 5 and its comparison with the gen
A:Reference number: A39449; MUID:92087470; PMID:1127603

A:Accession: G39449
A:Molecule type: DNA

A:Residues: 1-368 <CHR>

A:Cross-references: GB:M73260; GB:M29978; NID:G209842; PIDN:AAA96409.1; PID:G209846

C:Superfamily: adenovirus minor core protein pV
C:Keywords: core protein; late protein

Query Match 72.3%; Score 34; DB 1; Length 368;
Best Local Similarity 62.5%; Pred. No. 49;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVW 8
DB 59 APRRVQM 66

RESULT 11
FOADM2

minor core protein pV - human adenovirus 2

C:Species: Mastadenovirus h2 (human adenovirus 2)

A:Note: host Homo sapiens (man)

C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 04-Mar-1994
C:Accession: A03837

R:Allestree, P.; Akusjaervi, G.; Lager, M.; Yeh-Kai, L.; Pettersson, U.
J. Biol. Chem. 259, 13980-13985, 1984

A:Title: Genes encoding the core proteins of adenovirus type 2.
A:Reference number: A03837; MUID:85054835; PMID:6094534

A:Accession: A03837
A:Molecule type: DNA

A:Residues: 1-369 <ABE>

C:Superfamily: adenovirus minor core protein pV
C:Keywords: core protein; late protein

Query Match 72.3%; Score 34; DB 1; Length 369;
Best Local Similarity 62.5%; Pred. No. 49;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVW 8
DB 60 APRRVQM 67

RESULT 12

T17882

hypothetical protein a381R - Chlorella virus PBCV-1

C:Species: Chlorella virus PBCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T17882
R:Graves, M.V.; Van Etten, J.L.

submitted to the EMBL Data Library, May 1999

A:Reference number: Z18806

A:Accession: T17882

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-85 <GRA>

A:Cross-references: EMBL:U42580; NID:G4028896; PIDN:AAC96749.1

A:Experimental source: specific host Chlorella strain NC6A

C:Genetics:

A:Note: a381R

Query Match 70.2%; Score 33; DB 2; Length 85;
Best Local Similarity 75.0%; Pred. No. 19;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVW 8
DB 66 APRKNVW 73

RESULT 13

T30991

hypothetical protein C01G5.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 04-Mar-2000

C:Accession: T30991

R:Bradsaw, H.; Stellyes, L.

submitted to the EMBL Data Library, August 1999

A:Description: The sequence of C. elegans cosmid C01G5.

A:Reference number: Z20956

A:Accession: T30991

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-164

A:Cross-references: EMBL:U50068; PIDN:AAB37737.1

A:Experimental source: strain Bristol NZ

C:Genetics:

A:Map position: IV

A:introns: 26/1; 70/1; 125/1

A:Note: C01G5.7

C:Superfamily: Caenorhabditis elegans hypothetical protein C01G5.7

Query Match 70.2%; Score 33; DB 2; Length 164;
Best Local Similarity 83.3%; Pred. No. 36;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 RKXVW 8
DB 155 KXVW 160

RESULT 14

RGH0UA

regenerating islet lectin 1-alpha precursor [validated] - human

N:Alternate names: lithostathine; pancreatic thread protein (PTP); reg I protein; regI-a

N:Contains: pancreatic stone protein (PSP)

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1990 #sequence_revision 03-Aug-1995 #text_change 08-Dec-2000

C:Accession: A35197; B28351; S12950; S02767; S02419; S00113; S01471; A25246

R:Watanabe, T.; Ionomura, H.; Terazono, K.; Yamamoto, H.; Okamoto, H.

J. Biol. Chem. 265, 7432-7439, 1990

A:Title: Complete nucleotide sequence of human reg gene and its expression in normal and

product of the gene.

A:Reference number: A35197; MUID:90237042; PMID:232435

A:Accession: A35197

A:Molecule type: DNA

A:Residues: 1-166 <WAT>

A:Cross-references: GB:U05412

R:Terazono, K.; Yamamoto, H.; Takasawa, S.; Shiga, K.; Yonemura, Y.; Tochino, Y.; Okamoto

J. Biol. Chem. 263, 2111-2114, 1988

A:Title: A novel gene activated in regenerating islets.

A:Reference number: A92704; MUID:88115343; PMID:2963000

A:Accession: B28351

A:Molecule type: mRNA

A/Residues: 1-166 <TER>
 A/Cross-references: GB:M18963; NID:9190978; PIDN:AAA56558.1; PID:9190979
 R/Itoh, T.; Tsumaki, H.; Kato, T.; Teraoka, H.; Matsumoto, K.; Yoshida, N.; Teraono, K.
 FEBS Lett. 272, 85-88, 1990
 A/Title: Isolation and characterization of human reg protein produced in Saccharomyces
 A/Reference number: S12950; MUID:91032149; PMID:2226837
 A/Accession: S12950
 A/Molecule type: protein
 A/Residues: 23-52;160-166 <ITD>
 A/Note: sequence determined from protein isolated after human cDNA sequence was cloned
 R/de Caro, A.M.; Adrich, Z.; Fournet, B.; Capon, C.; Bonicel, J.J.; de Caro, J.D.; Rover
 Biochim. Biophys. Acta 994, 281-284, 1989
 A/Title: N-terminal sequence extension in the glycosylated forms of human pancreatic stc
 A/Reference number: S02767; MUID:89150292; PMID:2493268
 A/Accession: S02767
 A/Molecule type: protein
 A/Residues: 23-47 <DEC>
 R/Rouimi, P.; de Caro, J.J.; Bonicel, J.J.; Rovery, M.; de Caro, A.
 FEBS Lett. 229, 171-174, 1988
 A/Title: The disulfide bridges of the immunoreactive forms of human pancreatic stone pro
 A/Reference number: S02419; MUID:88152214; PMID:3345835
 A/Accession: S02419
 A/Molecule type: protein
 A/Residues: 63-72;125-139;150-157;160-166 <ROU>
 A/Note: disulfide bonds
 R/de Caro, A.M.; Bonicel, J.J.; Rouimi, P.; de Caro, J.D.; Sarles, H.; Rovery, M.
 Eur. J. Biochem. 168, 201-207, 1987
 A/Title: Complete amino acid sequence of an immunoreactive form of human pancreatic ston
 A/Reference number: S00113; MUID:88029417; PMID:3665916
 A/Accession: S00113
 A/Molecule type: protein
 A/Residues: 34-166 <DEI>
 R/Rouimi, P.; Bonicel, J.J.; Rovery, M.; de Caro, A.
 FEBS Lett. 216, 195-199, 1987
 A/Title: Cleavage of the Arg-Ile bond in the native polypeptide chain of human pancreati
 A/Reference number: S01471; MUID:87219142; PMID:3106036
 A/Accession: S01471
 A/Molecule type: protein
 A/Residues: 33-48 <RO2>
 R/Montalto, G.; Bonicel, J.J.; Multigner, L.; Rovery, M.; Sarles, H.; De Caro, A.
 Biochem. J. 238, 227-233, 1986
 A/Title: Partial amino acid sequence of human pancreatic stone protein, a novel pancreat
 A/Reference number: A25246; MUID:87093950; PMID:3541906
 A/Accession: A25246
 A/Molecule type: protein
 A/Residues: 34-73, 'X', '75-87', 'R', '89-98' <MON>
 C/Comment: This protein is found in pancreatic calculi of mammals. The tryptic-like clea
 C/Comment: Intact regenerating islet lectin 1-alpha, lithostathine, inhibits the growth
 C/Genetics:
 A/Gene: GDB:RG13A; REG
 A/Cross-references: GDB:132455; OMIM:167770
 A/Map position: 2p12-2p12
 A/Introns: 22/1; 61/3; 107/3; 145/1
 C/Superfamily: tetraneurin; C-type lectin homology
 C/Keywords: glycoprotein; lectin; pancreas; pyroglyutamic acid
 F/1-22/Domain: signal sequence #status predicted <SIG>
 F/23-166/Product: regenerating islet lectin 1-alpha #status experimental <MAT>
 F/34-166/Product: pancreatic stone protein #status experimental <MAT>
 F/36-166/Domain: C-type lectin homology <LCH>
 F/23/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimen
 F/27/Binding site: carbonylase (Thr) (covalent) #status experimental
 F/33-34/Cleavage site: Arg-Ile (tryptin) #status experimental
 F/36-47, 64-162, 137-154/Disulfide bonds: #status experimental

Query Match 70.2%; Score 33; DB 1; Length 166;
 Best Local Similarity 71.4%; Pred. No. 36;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRKXRVW 8
 DB 105 PKXNRW 111

RESULT 15
 RGHUB
 regenerating islet lectin 1-beta precursor - human
 N/Alternate names: reg-related protein; reg1-beta protein
 N/Contents: pancreatic stone protein (PSP)
 C/Species: Homo sapiens (man)
 C/Date: 22-Nov-1993 #sequence revision 03-Aug-1995 #text_change 16-Jun-2000
 R/Bartoli, C.; Garib, B.; Giorgi, D.; Sansonetti, A.; Dagnon, J.C.; Berge-Jefranc, J.L
 FEBS Lett. 327, 289-293, 1993
 A/Title: A gene homologous to the reg gene is expressed in the human pancreas.
 A/Reference number: S34591; MUID:93351647; PMID:8348956
 A/Accession: S34591
 A/Molecule type: DNA
 A/Residues: 1-166 <BAR>
 A/Cross-references: GB:L08010; NID:9307368; PIDN:AAA18204.1; PID:9487726
 A/Note: this gene appears to be expressed in pancreas and liver
 R/Morimoto, S.; Watanabe, T.; Unno, M.; Nakagawara, K.; Suzuki, Y.; Miyashita, H.; Yone
 Biochim. Biophys. Acta 1217, 199-202, 1994
 A/Title: Isolation, structural determination and expression of a novel reg gene, human r
 A/Reference number: S42729; MUID:94153997; PMID:8110835
 A/Accession: S42729
 A/Molecule type: mRNA
 A/Residues: 1-166 <MOR>
 A/Cross-references: GB:D16816; NID:9474305; PIDN:BA04091.1; PID:9474306
 A/Accession: A44712
 A/Molecule type: DNA
 A/Residues: 1-166 <MO2>
 A/Cross-references: GB:D17291; NID:9474307; PIDN:BA04124.1; PID:9474308
 C/Comment: This protein is found in pancreatic calculi of mammals. The tryptic-like clea
 C/Genetics:
 A/Gene: GDB:RG1B; REG1
 A/Cross-references: GDB:342079
 A/Map position: 2p12-2p12
 A/Introns: 22/1; 61/3; 107/3; 145/1
 C/Superfamily: tetraneurin; C-type lectin homology
 C/Keywords: glycoprotein; lectin; pancreas; pyroglyutamic acid
 F/1-22/Domain: signal sequence #status predicted <SIG>
 F/23-166/Product: regenerating islet lectin 1-beta #status predicted <MAT>
 F/34-166/Product: pancreatic stone protein #status predicted <MAT>
 F/36-166/Domain: C-type lectin homology <LCH>
 F/23/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
 F/27/Binding site: carbonylase (Thr) (covalent) #status predicted
 F/33-34/Cleavage site: Arg-Ile (tryptin) #status predicted
 F/36-47, 64-162, 137-154/Disulfide bonds: #status predicted

Query Match 70.2%; Score 33; DB 1; Length 166;
 Best Local Similarity 71.4%; Pred. No. 36;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRKXRVW 8
 DB 105 PKXNRW 111

Search completed: July 30, 2003, 16:31:22
 Job time: 8.09524 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd..

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:09:50 ; Search time 15.7619 Seconds
(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-8
Perfect score: 47
Sequence: 1 APRKXVW 8

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	100.0	8	21	AAV88343
2	47	100.0	18	17	AAE91193
3	47	100.0	51	17	AAE91194
4	47	100.0	52	16	AAE87909
5	47	100.0	689	21	AAI49270
6	47	100.0	704	22	AAW51535
7	47	100.0	708	17	AAW09343
8	47	100.0	708	19	AAW57318
9	47	100.0	708	20	AAW86022

10	47	100.0	708	21	AAE08183	Amino acid sequenc
11	47	100.0	708	22	AAE64827	Chronic hepatitis
12	47	100.0	708	22	AAE02342	Bovine lactoferrin
13	38	80.9	685	12	AAE11663	Partial porcine la
14	38	80.9	685	12	AAE11664	Partial porcine la
15	38	80.9	703	17	AAW09344	Porcine lactoferr
16	38	80.9	703	19	AAW57319	Porcine lactoferr
17	38	80.9	703	20	AAW86023	Porcine lactoferr
18	38	80.9	703	21	AAE08184	Amino acid sequenc
19	38	80.9	703	22	AAE02343	Porcine lactoferr
20	37	78.7	137	22	AAE00538	Human polyepitide
21	37	78.7	333	22	AAW00957	Human bone marrow
22	37	78.7	380	23	AAE90775	Human Tumour Endot
23	37	78.7	380	24	AAE54482	Human normal endot
24	37	78.7	560	23	AAE51300	Human MDR SEQ ID
25	36	76.6	708	20	AAE80774	Goat lactoferrin-as
26	36	76.6	708	20	AAE80775	Goat lactoferrin-as
27	35	74.5	61	22	AAE57597	Propionibacterium
28	35	74.5	404	22	AAE82055	S. epidermidis ope
29	35	74.5	406	23	AAE39239	Staphylococcus epi
30	35	74.5	807	19	AAW70465	Sindbis virus nsP2
31	34	72.3	98	23	AAE08720	Human ORFX protein
32	34	72.3	98	23	AAE08720	Human Apoptosis re
33	34	72.3	170	22	AAE60528	Propionibacterium
34	34	72.3	305	21	AAE56616	Human prostate can
35	34	72.3	337	22	AAE60661	Human apoptosis-re
36	34	72.3	356	24	AAE56645	Chimpanzee C68 ade
37	34	72.3	389	22	AAU14209	Human novel protei
38	34	72.3	439	21	AAE84905	A human prolifera
39	34	72.3	439	22	AAE93327	Human polypeptide,
40	34	72.3	439	22	AAU14445	Human novel protei
41	34	72.3	439	23	AAE05543	Breast cancer-asso
42	34	72.3	439	24	AAE47559	Breast cancer asso
43	34	72.3	439	24	AAE37023	Human breast cance
44	34	72.3	11300	23	AAE84277	Adenovirus Ad1ip a
45	34	72.3	19938	24	AAE98398	Streptomyces virid

ALIGNMENTS

RESULT 1
AAV88343
ID AAV88343 standard; peptide; 8 AA.
XX
AC AAV88343;
XX
XX
14-JUN-2000 (first entry)
XX
DT
XX
DE N-terminal peptide derived bovine whey.
XX
XX Peptide production; biological fluid; milk; whey; blood; bovine;
KW antibacterial peptide; lactoferrin; antiviral; antitumour activity.
XX
OS Bos sp.
XX
XX WO200015655-A1.
XX
XX 23-MAR-2000.
XX
XX 15-SEP-1999; 99WO-EP07002.
XX
XX 15-SEP-1998; 98EP-0203107.
XX
XX 08-JUN-1999; 99EP-0201815.
XX
XX (NIZO-) NIZO FOOD RES.
XX
XX Visser S, Recio I;
XX
XX WPI; 2000-271377/23.
XX
XX Novel process for producing peptides with e.g. antimicrobial activity
XX from biological fluids such as milk, whey or blood comprises contacting

PT fluid with chromatographic medium to adsorb peptide domain of interest
PS Claim 14; Page 22; 41pp; English.
XX
XX
CC This sequence represents an N-terminal peptide derived from bovine whey.
CC The peptide is an example of a peptide with antibacterial activity that
CC can be produced by the process of the invention. The invention relates to
CC a process for producing peptides from biological fluids. The process
CC comprises chromatography of the biological fluid, in situ hydrolysis of
CC selectively bound peptides, washing to remove unbound peptide, and
CC elution of the peptides of interest. The process is used for producing
CC peptides from biological fluids, such as milk, whey or blood. For
CC example, the process can be used to produce antibacterial peptides
CC derived from lactoferrin, using cheese whey as a starting material. The
CC peptides obtained have preferably antimicrobial and/or antiviral and/or
CC antitumor activity. The process of the invention is relatively simple
CC and generally economically and technically more attractive than those
CC methods previously used. The method provides high yield peptides with a
CC selected activity of interest without the need for intermediate
CC purification of the precursor protein.
SQ Sequence 8 AA;
Query Match 100.0%; Score 47; DB 21; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 APRKNVRW 8
DB 1 APRKNVRW 8
RESULT 2
ID AAR91193 standard; peptide; 18 AA.
XX AAR91193;
AC AAR91193;
XX
DT 06-SEP-1996 (first entry)
XX
DE Lactoferrin decomposition peptide.
XX
KM immunosactivator; antiviral; cytomegalovirus; cosmetic; food; feed;
KM lymphocyte blastogenesis.
XX
OS Synthetic.
XX
PN JP08073499-A.
XX
PD 19-MAR-1996.
XX
PF 01-SEP-1994; 94JP-0232026.
XX
PR 01-SEP-1994; 94JP-0232026.
XX
PA (SNOW) SNOW BRAND MILK PROD CO LTD.
XX
DR WPT; 1996-205535/21.
XX
PT New peptide(s) derived from human lactoferrin - are useful as
PT immunoactivators, esp. for preventing infection by cytomegalovirus
PS Claim 4; Page 2; 11pp; Japanese.
XX
XX The present sequence is one of four new peptides (see AAR91191 -
CC AAR91193) obtained by enzymatically decomposing lactoferrin using
CC protease. The peptides are immunoactivators which induce
CC blastogenesis of lymphocytes and are particularly useful for
CC treating cytomegalovirus infection. They can be used in drugs and
CC cosmetics and can be added to foods and feeds.
SQ Sequence 18 AA;

Query Match 100.0%; Score 47; DB 17; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 APRKNVRW 8
DB 1 APRKNVRW 8
RESULT 3
ID AAR91194 standard; peptide; 51 AA.
XX AAR91194
AC AAR91194;
XX
DT 06-SEP-1996 (first entry)
XX
DE Lactoferrin decomposition peptide.
XX
KM immunosactivator; antiviral; cytomegalovirus; cosmetic; food; feed;
KM lymphocyte blastogenesis.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Disulfide-bond 9..45 "optionally this bond may be reduced"
FT Disulfide-bond 19..36 "optionally this bond may be reduced"
FT /note="optionally this bond may be reduced"
XX
PN JP08073499-A.
XX
PD 19-MAR-1996.
XX
PF 01-SEP-1994; 94JP-0232026.
XX
PR 01-SEP-1994; 94JP-0232026.
XX
PA (SNOW) SNOW BRAND MILK PROD CO LTD.
XX
DR WPT; 1996-205535/21.
XX
PT New peptide(s) derived from human lactoferrin - are useful as
PT immunoactivators, esp. for preventing infection by cytomegalovirus
PS Claim 5; Page 2; 11pp; Japanese.
XX
XX The present sequence is one of four new peptides (see AAR91191 -
CC AAR91193) obtained by enzymatically decomposing lactoferrin using
CC protease. The peptides are immunoactivators which induce
CC blastogenesis of lymphocytes and are particularly useful for
CC treating cytomegalovirus infection. They can be used in drugs and
CC cosmetics and can be added to foods and feeds.
SQ Sequence 51 AA;
Query Match 100.0%; Score 47; DB 17; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 APRKNVRW 8
DB 1 APRKNVRW 8
RESULT 4
ID AAR87909 standard; peptide; 52 AA.
XX AAR87909;
AC AAR87909;
XX
DT 01-MAR-1996 (first entry)


```

XX DE Bovine lactoferrin (1-52).
XX XX antiviral; lactoferrin;
XX OS Synthetic.
XX Key Location/Qualifiers
XX FH Disulfide-bond 9...45
XX FT /note= "each Cys residue may be in reduced form or may
XX FT form a disulphide bond with another Cys residue.
XX FT A 9-45 disulphide bond is exemplified"
XX FT Disulfide-bond 19...36
XX FT /note= "each Cys residue may be in reduced form or may
XX FT form a disulphide bond with another Cys residue.
XX FT A 19-36 disulphide bond is exemplified"
XX PM JF07069915-A.
XX PD 14-MAR-1995.
XX PF 02-SEP-1993; 93JP-0240284.
XX PR 02-SEP-1993; 93JP-0240284.
XX PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
XX DR WPI; 1995-144726/19.
XX XX Inhibitor against viral infection and proliferation - contains
XX PT peptide having sequence from lactoferrin
XX PS Claim 5; Page 2; 10pp; Japanese.
XX CC The sequence is one of six peptides disclosed as having inhibitory
XX CC effect against viral infection. The peptides are derived from
XX CC lactoferrin. Their activity is demonstrated against cytomegalovirus.
XX SQ Sequence 52 AA;

Query Match 100.0%; Score 47; DB 16; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
DB 1 APRKNVRW 8

RESULT 5
AAAY49270
ID AAAY49270 standard; protein; 689 AA.
XX AC AAAY49270;
XX DT 07-FEB-2000 (first entry)
XX DE Lactoferrin sequence.
XX KM Adhesion inhibition; enteropathogenic; Escherichia coli; lactoferrin;
XX KM medicament; drink; feed; food poisoning.
XX OS Unidentified.
XX PN JF11292789-A.
XX PD 26-OCT-1999.
XX PF 03-APR-1998; 98JP-0107167.
XX PR 03-APR-1998; 98JP-0107167.
XX PA (MEIP ) MEIJI MILK PROD CO LTD.
XX PA

```

```

XX XX WPI; 2000-018674/02.
XX XX Adhesion inhibitory compositions of enteropathogenic E. coli to cells -
XX PT used in medicines and foodstuffs
XX PS Disclosure; Fig 4; 7pp; Japanese.
XX CC The invention provides an adhesion inhibitory composition of
XX CC enteropathogenic Escherichia coli to cells. The composition contains a
XX CC lactoferrin degraded matter. The composition is useful as a medicament,
XX CC a drink and food or a feed. The composition is effective for prevention
XX CC and/or treatment of infectious food poisoning caused by enteropathogenic
XX CC E. coli.
XX SQ Sequence 689 AA;

Query Match 100.0%; Score 47; DB 21; Length 689;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
DB 1 APRKNVRW 8

RESULT 6
AAM51535
ID AAM51535 standard; Protein; 704 AA.
XX AC AAM51535;
XX DT 07-JAN-2002 (first entry)
XX DE Recombinant lactoferrin.
XX KM Bovine; recombinant; lactoferrin; pWE; Candida methyllica SKL60.
XX OS Bos sp.
XX PN KR2001039000-A.
XX PD 15-MAY-2001.
XX PF 28-OCT-1999; 99KR-0047203.
XX PR 28-OCT-1999; 99KR-0047203.
XX PA (EASY-) EASY BIO SYSTEM INC.
XX PA (SUNG/) SUNG C G.
XX PI Min DG, Mun TH, Sung CG;
XX DR WPI; 2001-594660/67.
XX DR N-PSDB; AA169871.
XX PT Manufacturing Candida methyllica SKL60 (Kctc 8968p) and recombinant
XX PT lactoferrin -
XX PS Disclosure; Fig 5; 15pp; Korean.
XX CC The invention relates to a method for manufacturing a recombinant yeast
XX CC strain for lactoferrin expression. A recombinant vector pWE containing
XX CC the lactoferrin gene is used to transform Candida methyllica SKL60,
XX CC which, in turn, produces recombinant lactoferrin at high yield.
XX CC The method of manufacturing a recombinant lactoferrin gene comprises:
XX CC (i) separating RNA from Korean cow, using antibacterial cDNA to make a
XX CC primer for amplifying antibacterial lactoferrin DNA, and inserting
XX CC the lactoferrin gene into vector pWE to make an expression vector pWE;
XX CC (ii) electrically shocking Candida methyllica KXL-6 to insert the
XX CC expression vector pWE into the bacteria;
XX CC (iii) pre-inducing the transformed strain with 10 % of glycerol and
XX CC inducing the transformant with 0.5-6 % of ethanol, and

```

CC (iv) cultivating the bacteria in pH 4.0-6.0 of a fermenter at 30 plusoc
 CC for 14-20 hours.
 CC The present sequence is the lactoferrin polypeptide.

XX Sequence 704 AA;

Query Match 100.0%; Score 47; DB 22; Length 704;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
 DB 20 APRKNVRW 27

RESULT 7
 ID AAM09343 standard; Protein; 708 AA.

XX AAM09343;

XX 25-MAR-2003 (updated)

DT 18-MAR-1997 (first entry)

XX Bovine lactoferrin.

XX Human; lactoferrin; iron-binding glycoprotein; milk; secretion; fungus;
 KM transferin; bactericidal activity; prostatic; expression system; primer;
 KM PCR; polymerase chain reaction; amplification; signal peptide; antiviral;
 KM alpha-amylinase; Aspergillus oryzae; nutrition; bovine.

XX Bos taurus.

XX US5571691-A.

XX 05-NOV-1996.

PF 28-OCT-1993; 93US-0145681.

XX 28-OCT-1993; 93US-0145681.

PR 05-MAY-1989; 89US-0348270.

PR 28-SEP-1989; 89US-0413880.

PR 24-APR-1992; 92US-0873304.

PR 27-OCT-1992; 92US-0967947.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Connely OM, Heaton DR, May GS, O'Malley BW;

DR MPI; 1996-505400/50.

XX N-PSDB; AAT48031.

XX New DNA encoding human lactoferrin, its natural alleles and
 PT substitution analogues - useful e.g. for preventing iron deficiency
 PT and as antiviral/antimicrobial agent

XX Disclosure; Column 31-34; 92pp; English.

XX This is the amino acid sequence of bovine lactoferrin, an iron-binding
 CC glycoprotein found in milk and other secretions and body fluids.

CC Fragments of the protein are also known to have biological activity e.g.
 CC the N-terminal portion of the protein has a bactericidal activity. The
 CC gene was used to construct a fusion protein in which the native
 CC lactoferrin signal peptide was replaced by the alpha-amylinase II signal
 CC peptide. The novel construct was then expressed in Aspergillus oryzae.
 CC The protein can be used for antibacterial and antiviral activities as
 CC well as an iron-carrying protein for nutritional or therapeutic
 CC applications.

CC (Updated on 25-MAR-2003 to correct PF field.)

XX Sequence 708 AA;

XX Query Match

100.0%; Score 47; DB 17; Length 708;

Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
 DB 20 APRKNVRW 27

RESULT 8
 ID AAM57318 standard; Protein; 708 AA.

XX AAM57318;

XX 10-AUG-1998 (first entry)

XX Bovine lactoferrin.

XX Bovine; lactoferrin; recombinant; therapeutic; nutritional; iron;
 KM Fe binding site; bacteria; bactericidal; milk.

XX Bos taurus.

XX US5766939-A.

XX 16-JUN-1998.

XX 30-MAY-1995; 95US-0453703.

XX 28-OCT-1993; 93US-0145681.

XX 05-MAY-1989; 89US-0348270.

XX 24-APR-1992; 92US-0873304.

XX 27-OCT-1992; 92US-0967947.

XX 30-MAY-1995; 95US-0453703.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Connely OM, Heaton DR, May GS, O'Malley BW;

XX MPI; 1998-361705/31.

XX N-PSDB; AAV30771.

XX Plasmids containing human lactoferrin DNA - for recombinant
 PT production of the enzyme, especially fragments having bactericidal
 PT activity

XX Example 13; Fig 14; 92pp; English.

XX The present sequence represents bovine lactoferrin. A plasmid has been
 CC developed which is suitable for the expression of a human lactoferrin,
 CC or an iron-binding lobe of lactoferrin, in a transformed prokaryotic
 CC host cell. The plasmid comprises a DNA sequence encoding a naturally
 CC occurring human lactoferrin protein or an iron-binding lobe of
 CC lactoferrin and further comprises transcriptional and translational
 CC regulatory elements capable of regulating the expression of the
 CC lactoferrin-encoding DNA sequence in the transformed host cell. The
 CC plasmid is useful for producing recombinant human lactoferrin proteins
 CC in bacteria. Lactoferrin is a 78 kDa iron-binding glycoprotein found
 CC in milk and other secretory fluids. It is involved in iron transfer
 CC and delivery in mammals. It has been implicated as a resistance
 CC factor in suckled new born infants against enteric infections; the
 CC iron binding activity thought to be responsible for bactericidal action
 CC as the bound iron deprives microorganisms of the metal, and interferes
 CC with their growth and reproduction. The protein can be used for
 CC nutritional and therapeutic applications. It has not been previously
 CC possible to efficiently produce lactoferrin recombinantly.

XX Sequence 708 AA;

XX Query Match

100.0%; Score 47; DB 19; Length 708;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVW 8
 DB 20 APRKNVW 27

RESULT 9
 ID AAW86022 standard; Protein; 708 AA.
 AAW86022;

AC AAW86022;
 DT 20-MAR-2003 (updated)
 DT 22-FEB-1999 (first entry)

DE Bovine lactoferrin (hLF) polypeptide.

KW Lactoferrin; hLF; recombinant; iron-binding; bovine; porcine; eyedrop;
 KW Aspergillus; animal food; therapeutic additive; iron transport; human;
 KW virucidal; bactericidal; contact lens; skin care; eardrop; mouthwash;
 KW chewing gum; toothpaste; preservative; anti-infection; nutrition.

OS Bos taurus.

XX US5849881-A.

XX US5849881-A.

PD 15-DEC-1998.

XX 30-MAY-1995; 95US-0456106.

XX 28-OCT-1993; 93US-0145681.

XX 05-MAY-1989; 89US-0348270.

PR 24-APR-1992; 92US-0873304.

PR 27-OCT-1992; 92US-0967947.

PR 30-MAY-1995; 95US-0456106.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Conneely OM, Headon DR, May GS, O'Malley BW;

XX WPI; 1999-063808/06.

DR N-PSDB; AAV80376.

XX Recombinant lactoferrin - and lactoferrin fragments

XX Example 13; Columns 33-38; 93pp; English.

XX This represents a bovine lactoferrin polypeptide. The invention relates

CC to production of recombinant lactoferrin using cDNA sequences of human,

CC bovine and porcine lactoferrin. Regions of the cDNA such as the iron-

CC binding sites can be used to produce recombinant lactoferrin, especially

CC human lactoferrin (hLF) polypeptide product. Aspergillus host cell

CC transformed with an expression vector comprising a hLF DNA sequence, a

CC promoter, a translation initiation sequence and transcription and

CC translation termination sequences can be used for the production of the

CC recombinant hLF polypeptide. Recombinant lactoferrin polypeptides can be

CC used in human and animal foods, as therapeutic additives to enhance iron

CC transport and delivery, and for their virucidal and bactericidal

CC properties, as additive for eyedrops, contact lens care solutions, skin

CC care products, eardrops, mouthwashes, chewing gum and toothpaste, as

CC preservatives for the above products, as therapeutic anti-infection

CC agents, as nutritional supplements and as sources of amino acids.

CC (Updated on 20-MAR-2003 to correct PA field.)

CC Sequence 708 AA;

XX Query Match 100.0%; Score 47; DB 20; Length 708;

XX Best Local Similarity 100.0%; Pred. No. 2.7;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVW 8

DB 20 APRKNVW 27

RESULT 10
 ID AAB08183 standard; Protein; 708 AA.

AC AAB08183;
 DT 04-DEC-2000 (first entry)

DE Amino acid sequence of a bovine lactoferrin polypeptide.

KW Lactoferrin; plasmid; Aspergillus host cell; iron transport; eye drop;
 KW iron delivery; virucidal; bactericidal; additive; eye care solution;
 KW skin care; ear drop; mouth wash; chewing gum; tooth paste;
 KW iron carrier protein; bacteriostatic; bactericide; iron deficiency;
 KW disease resistance; infection.

OS Bos taurus.

XX US5100054-A.

XX US5100054-A.

PD 08-AUG-2000.

XX 30-MAY-1995; 95US-0456108.

XX 28-OCT-1993; 93US-0145681.

XX 05-MAY-1989; 89US-0348270.

PR 24-APR-1992; 92US-0873304.

PR 27-OCT-1992; 92US-0967947.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Headon DR, O'Malley BW, May GS, Conneely OM;

XX WPI; 2000-531666/48.

DR N-PSDB; AAB63789.

XX Recombinant expression vector for producing lactoferrin protein useful

PT as therapeutic additive comprises nucleotide sequence encoding allelic

PT product, iron binding lobe or antimicrobial peptide of lactoferrin -

XX product, iron binding lobe or antimicrobial peptide of lactoferrin -

XX Disclosure; Fig 14C; 93pp; English.

XX The present sequence represents a bovine lactoferrin polypeptide. The

CC lactoferrin polynucleotide sequence is used to produce novel plasmids.

CC These plasmids have a transcription unit comprising in operable 5' to 3'

CC linkage: a promoter and nucleotide sequence permitting translation

CC initiation in an Aspergillus host cell, a lactoferrin-encoding nucleotide

CC sequence and nucleotide sequences permitting translation and

CC transcription termination in a Aspergillus host cell. The vector allows

CC expression of human lactoferrin in and subsequent processing by a

CC transformed fungal host cell. The lactoferrin is useful as a therapeutic

CC additive to enhance iron transport and delivery, for virucidal and

CC bactericidal qualities, as additives for eye drops, contact lens and

CC other eye care solutions, topical skin care products, ear drops, mouth

CC washes, chewing gum and tooth paste. The iron binding lactoferrin

CC polypeptides are useful as iron carrier proteins for nutritional and

CC therapeutic uses, bacteriostats and bactericides. The incorporation of

CC a human lactoferrin gene and expression in the milk of animals can

CC combat an iron deficiency typical in piglets. The inclusion of the

CC human lactoferrin gene with expression improves an animal's disease

CC resistance to bacterial and viral infection. The tissue specific

CC expression of human lactoferrin in mammary glands, for instance,

CC imparts the bactericidal and virucidal benefit of the expressed gene to

CC young feeding on the milk and provides a production means for the

CC secreted protein for therapeutic use.

XX Sequence 708 AA;

XX Query Match 100.0%; Score 47; DB 21; Length 708;

XX Best Local Similarity 100.0%; Pred. No. 2.7;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKVRM 8
 |||||
 DB 20 APRKVRM 27

RESULT 11

ID AAG64827 standard; protein; 708 AA.

AC AAG64827;

DT 20-SEP-2001 (first entry)

DE Chronic hepatitis treatment related protein SEQ ID NO: 8.

KM Chronic hepatitis; viral antigenic protein; hepatitis C; hepatitis B.

OS Bos taurus.

PM WO200147545-A1.

PD 05-JUL-2001.

PF 28-DEC-2000; 2000WO-JP09393.

PR 28-DEC-1999; 99JP-0374087.

PA (SUMU) SUMITOMO PHARM CO LTD.

PI Tohdoh N, Murata M, Enjoji T;

DR WPI; 2001-425585/45.

PT Treatment and prevention of chronic hepatitis

PS Example 1; Page 74-78; 128pp; Japanese.

CC The present invention describes a method of preventing and treating

CC chronic hepatitis, involving administering an oligopeptide which

CC (a) has binding affinity towards the viral antigenic protein;

CC (b) inhibits binding affinity of the virus towards the receptor protein

CC of the target cell and

CC (c) has analogy with the receptor protein at the amino acid level. This

CC can be used to prevent and treat hepatitis B and C. The present sequence

CC is a protein described in the exemplification of the invention.

SQ Sequence 708 AA;

QY 1 APRKVRM 8
 |||||
 DB 20 APRKVRM 27

RESULT 12

ID AAE02342 standard; Protein; 708 AA.

AC AAE02342;

DT 10-AUG-2001 (first entry)

DE Bovine lactoferrin (LF).

KM Bovine; lactoferrin; LF; therapeutic; nutritional; iron transport;

KM virucidal; bactericidal; animal food; iron-binding glycoprotein.

OS Bos taurus.

PM US6228614-B1.

XX 08-MAY-2001.
 PD 09-MAR-1999; 99US-0265577.
 PF 28-OCT-1993; 93US-0145681.

PR 05-MAY-1989; 89US-0348270.

PR 30-MAY-1995; 95US-0456108.

PR 24-APR-1992; 92US-0873304.

PR 27-OCT-1992; 92US-0967947.

PA (BAYU) BAYLOR COLLEGE MEDICINE.

PI Connely OM, Headon DR, O'malley BW, May GS;

PM WPI; 2001-342673/36.

PD N-PSDB; AAD06282.

PF Producing recombinant human lactoferrin useful for therapeutic or

PT nutritional applications, comprises transforming a eukaryotic cells

PT with a vector having a DNA that encodes and permits expression of

PT lactoferrin in the cells

PS Example 13; Fig 14B; 92pp; English.

CC The present invention relates to a method for expressing human

CC lactoferrin in an eukaryotic cell. The method comprises introducing

CC a plasmid comprising DNA encoding a human lactoferrin protein and

CC regulatory elements necessary for the expression of DNA in the cell.

CC The method is used for preparing recombinant lactoferrin, making

CC available a source of protein for therapeutic and nutritional

CC applications, e.g. as therapeutic additives to enhance iron transport

CC and delivery and for virucidal or bactericidal qualities or as human

CC or animal food. The present method is an efficient and economical way

CC to produce human lactoferrin.

CC The present sequence is bovine lactoferrin (LF). LF is an

CC iron-binding glycoprotein found in milk, other secretions and body

CC fluids.

SQ Sequence 708 AA;

QY 1 APRKVRM 8
 |||||
 DB 20 APRKVRM 27

DE 04-JUL-1991 (first entry)

DE Partial porcine lactoferrin protein.

KM Porcine lactoferrin; iron transport; iron binding.

OS Sus scrofa.

PM WO9105045-A.

PD 18-APR-1991.

PF 14-SEP-1990; 90WO-US05245.

XX PR 28-SEP-1989; 89US-0413880.
 XX XX (GRAN-) GRANADA BIOSCIENCES.
 XX PA Headon DR, Conneely OM, O'Malley BW;
 XX PI WPI; 1991-132852/18.
 XX DR N-PSDB; AAQ11559.
 XX PT New cDNA and amino acid sequence of porcine lactoferrin - has
 PT therapeutic and nutritional applications, especially iron
 PT transport and delivery
 XX XX Disclosure; fig 1; 19pp; English.
 XX CC This sequence comprises a portion of the porcine lactoferrin protein
 CC (most of signal peptide and part of mature protein incl. two iron
 CC binding domains). This protein, or a constituent polypeptide, contg.
 CC at least one of the iron binding domains, is useful for eg improving
 CC an animal's resistance to bacterial and viral infection. It can be used
 CC in a variety of prodcs. incl. feed, therapeutic additives and additives
 CC for eye drops, contact lens solns., etc. The partial porcine lacto-
 CC ferrin signal peptide can be replaced by a human lactoferrin signal
 CC peptide (see AAR11664).
 XX XX
 XX SQ Sequence 685 AA;
 SO Query Match 80.9%; Score 38; DB 12; Length 685;
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 APRKNVNM 8
 ||:||||
 20 APRKGVNM 27
 Db
 RESULT 14
 AAR11664
 ID AAR11664 standard; Protein; 685 AA.
 XX AC AAR11664;
 XX XX
 XX DT 04-JUL-1991 (first entry)
 XX XX
 XX DE Partial porcine/human lactoferrin fusion protein.
 XX XX
 XX KW Porcine lactoferrin; iron transport; iron binding.
 XX XX
 XX OS Sus scrofa.
 XX XX
 XX FH Key Location/Qualifiers
 FT Peptide 1..19
 FT /label= human lactoferrin signal peptide
 FT Protein 20..680
 FT /label= partial mature porcine lactoferrin
 XX XX
 XX PN MO9105045-A.
 XX PD 18-APR-1991.
 XX PF 14-SEP-1990; 90WC-US05245.
 XX PR 28-SEP-1989; 89US-0413880.
 XX XX
 XX PA (GRAN-) GRANADA BIOSCIENCES.
 XX PI Headon DR, Conneely OM, O'Malley BW;
 XX XX WPI; 1991-132852/18.
 XX DR N-PSDB; AAQ11560.
 XX PT New cDNA and amino acid sequence of porcine lactoferrin - has

PT therapeutic and nutritional applications, especially iron
 PT transport and delivery
 XX XX Disclosure; fig 2; 19pp; English.
 XX CC This sequence comprises a portion of the porcine lactoferrin prot-
 CC ein (incl. two iron binding domains). This protein, or a consti-
 CC uent polypeptide, contg. at least one of the iron binding domains,
 CC is useful for eg improving an animal's resistance to bacterial and
 CC viral infection. It can be used in a variety of prodcs. incl. feed,
 CC therapeutic additives and additives for eye drops, contact lens
 CC solns., etc.
 XX XX See also AAR11663.
 XX XX
 XX SQ Sequence 685 AA;
 SO Query Match 80.9%; Score 38; DB 12; Length 685;
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 APRKNVNM 8
 ||:||||
 20 APRKGVNM 27
 Db
 RESULT 15
 AAW09344
 ID AAW09344 standard; Protein; 703 AA.
 XX AC AAW09344;
 XX XX
 XX DT 25-MAR-2003 (updated)
 DT 18-MAR-1997 (first entry)
 XX XX
 XX DE Porcine lactoferrin.
 XX XX
 XX KW Human; lactoferrin; iron-binding glycoprotein; milk; secretion; fungus;
 KW transferrin; bactericidal activity; prostate; expression system; primer;
 KW PCR; polymerase chain reaction; amplification; signal peptide; antiviral;
 KW alpha-amylase; Aspergillus oryzae; nutrition; porcine.
 XX XX
 XX OS Sus scrofa.
 XX XX
 XX EN US5571691-A.
 XX PD 05-NOV-1996.
 XX XX
 XX PF 28-OCT-1993; 93US-0145681.
 XX PR 28-OCT-1993; 93US-0145681.
 PR 05-MAY-1989; 89US-0348270.
 PR 28-SEP-1989; 89US-0413880.
 PR 24-APR-1992; 92US-0873304.
 PR 27-OCT-1992; 92US-0967947.
 XX XX
 XX PA (BAYU) BAYLOR COLLEGE MEDICINE.
 XX PI Headon DR, May GS, O'Malley BW;
 XX XX WPI; 1996-505400/50.
 XX DR N-PSDB; AAT48032.
 XX PT New DNA encoding human lactoferrin, its natural alleles and
 PT substitution analogues - useful e.g. for preventing iron deficiency
 PT and as antiviral/antimicrobial agent
 XX XX
 XX PS Disclosure; Column 37-40; 92pp; English.
 XX XX
 XX CC This is the amino acid sequence of porcine lactoferrin, an iron-binding
 CC glycoprotein found in milk and other secretions and body fluids.
 CC Fragments of the protein are also known to have biological activity e.g.
 CC the N-terminal portion of the protein has a bactericidal activity. The
 CC gene was used to construct a fusion protein in which the native

CC lactoferrin signal peptide was replaced by the alpha-amylase II signal
CC peptide. The novel construct was then expressed in *Aspergillus oryzae*.
CC The protease can be used for antibacterial and antiviral activities as
CC well as an iron-carrying protein for nutritional or therapeutic
CC applications.
CC (Updated on 25-MAR-2003 to correct PF field.)
CC XX
SQ Sequence 703 AA;

Query Match 80.9%; Score 38; DB 17; Length 703;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKQVRW 8
||:||||
Db 20 APKKGVRRW 27

Search completed: July 30, 2003, 16:23:31
Job time : 16.7619 secs